

Communication

Machine Learning-Guided Prediction of Antigen-Reactive In Silico Clonotypes Based on Changes in Clonal Abundance through Bio-Panning

Duck Kyun Yoo ^{1,2,†} , Seung Ryul Lee ^{1,3,†}, Yushin Jung ⁴, Haejun Han ⁴, Hwa Kyoung Lee ^{1,2}, Jerome Han ^{1,2}, Soohyun Kim ^{1,3}, Jisu Chae ^{1,3}, Taehoon Ryu ^{4,*} and Junho Chung ^{1,2,3,*}

¹ Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul 03080, Korea; dk93js@gmail.com (D.K.Y.); seungryul93@gmail.com (S.R.L.); intia6@gmail.com (H.K.L.); wsukh23@gmail.com (J.H.); kimchii481@gmail.com (S.K.); jschaelab@gmail.com (J.C.)

² Department of Biomedical Science, Seoul National University College of Medicine, Seoul 03080, Korea

³ Cancer Research Institute, Seoul National University College of Medicine, Seoul 03080, Korea

⁴ Celemics, Inc., 131 Gasandigital 1-ro, Geumcheon-gu, Seoul 08506, Korea;

yushinjung@celemics.com (Y.J.); hhjunny@gmail.com (H.H.)

* Correspondence: taehoon.ryu@celemics.com (T.R.); jhchung@snu.ac.kr (J.C.);

Tel.: +82-2-6746-8067 (T.R.); +82-2-3668-7439 (J.C.)

† These authors contributed equally to this work.

Received: 31 December 2019; Accepted: 6 March 2020; Published: 8 March 2020



Abstract: c-Met is a promising target in cancer therapy for its intrinsic oncogenic properties. However, there are currently no c-Met-specific inhibitors available in the clinic. Antibodies blocking the interaction with its only known ligand, hepatocyte growth factor, and/or inducing receptor internalization have been clinically tested. To explore other therapeutic antibody mechanisms like Fc-mediated effector function, bispecific T cell engagement, and chimeric antigen T cell receptors, a diverse panel of antibodies is essential. We prepared a chicken immune scFv library, performed four rounds of bio-panning, obtained 641 clones using a high-throughput clonal retrieval system (TrueRepertoireTM, TR), and found 149 antigen-reactive scFv clones. We also prepared phagemid DNA before the start of bio-panning (round 0) and, after each round of bio-panning (round 1–4), performed next-generation sequencing of these five sets of phagemid DNA, and identified 860,207 HCDR3 clonotypes and 443,292 LCDR3 clonotypes along with their clonal abundance data. We then established a TR data set consisting of antigen reactivity for scFv clones found in TR analysis and the clonal abundance of their HCDR3 and LCDR3 clonotypes in five sets of phagemid DNA. Using the TR data set, a random forest machine learning algorithm was trained to predict the binding properties of in silico HCDR3 and LCDR3 clonotypes. Subsequently, we synthesized 40 HCDR3 and 40 LCDR3 clonotypes predicted to be antigen reactive (AR) and constructed a phage-displayed scFv library called the AR library. In parallel, we also prepared an antigen non-reactive (NR) library using 10 HCDR3 and 10 LCDR3 clonotypes predicted to be NR. After a single round of bio-panning, we screened 96 randomly-selected phage clones from the AR library and found out 14 AR scFv clones consisting of 5 HCDR3 and 11 LCDR3 AR clonotypes. We also screened 96 randomly-selected phage clones from the NR library, but did not identify any AR clones. In summary, machine learning algorithms can provide a method for identifying AR antibodies, which allows for the characterization of diverse antibody libraries inaccessible by traditional methods.

Keywords: c-Met; next-generation sequencing; phage display; machine learning; random forest; antibody discovery

1. Introduction

The mesenchymal-epithelial transition factor (c-Met) and its ligand hepatocyte growth factor (HGF) are well-known oncogenic drivers of tumorigenesis [1]. Numerous clinical observations have demonstrated that c-Met overexpression or gene alterations play a key role in both oncogenesis and the development of drug resistance across multiple cancer types [2–5]. Furthermore, recent research suggests that the HGF-c-Met axis limits the efficacy of cancer immunotherapy by modulating immune cell function and the expression of programmed cell death ligand 1 (PD-L1) [6–9]. Despite efforts to inhibit the HGF-c-Met axis including antibodies against c-Met or HGF, c-Met tyrosine kinase inhibitors, and more, no therapeutic agent specific to the HGF-c-Met axis is clinically available. Currently, two anti-HGF antibodies, including YYB-101 previously discovered by our group, are under clinical trials (NCT02499224) [10]. However, no antibodies are under development against c-Met after the failure of onartuzumab in clinical trials [11].

Based on rapid advances in next-generation sequencing (NGS) technology, various methodologies for analyzing NGS data have been developed to decode the antibody repertoire from diverse sources such as the natural B cell receptor of animals and humans as well as recombinant antibody libraries that can be synthetically designed and constructed [12–14]. Furthermore, combining surface display technology and NGS analysis offers synergistic advantages in identifying antigen-reactive clones *in silico* over the laborious *in vitro* screening process, which is frequently overwhelmed by dominant antibody clones [15]. Traditional bio-panning methodologies are biased towards the excessive enrichment of dominant clones with significant suppression of antibody diversity. Consequently, this approach could lead to the omission of potential antigen-reactive (AR) clones with low clonal abundance or their diminishment by unintended selective pressure.

Previously, our group analyzed the enrichment patterns of bio-panned clones by employing NGS technology to predict the antigen binding properties of antibody clones inside different clusters [16]. First, we tracked the clonal abundance of heavy chain complementarity region 3 (HCDR3) through multiple rounds of bio-panning with NGS analysis, and then applied clustering analysis to group HCDR3 clonotypes based on the enrichment pattern. As a result, different clusters (enriched, impoverished, and fluctuated) were generated with the enriched pattern cluster containing a higher frequency of AR scFv (single-chain variable fragment) clones than other clusters. However, due to limitations in retrieving the physical DNA of encoded scFv from a large, diverse number of clones, we were unable to sufficiently observe the binding properties of *in silico* scFv clones. Recently, we developed a laser and microchip-based high-throughput clonal retrieval system (TrueRepertoireTM, TR) for scFv DNA from the library [17], which allows a much higher number of scFv clones to be obtained and tested for antigen reactivity.

In this study, we established a phage-displayed chicken scFv library after immunization with recombinant c-Met. Four rounds of bio-panning were performed on antigen-conjugated magnetic beads. Through bio-panning, five sets of phagemid DNA (rounds 0–4) were obtained and subjected to NGS analysis using both HiSeq and MiSeq platforms. After the final round of bio-panning, scFv-displayed phage clones were obtained in a high-throughput manner using TR technology, and individual clone reactivity was evaluated by phage enzyme-linked immunosorbent assay (ELISA). From NGS data obtained using the HiSeq platform, HCDR3, and light chain complementarity region 3 (LCDR3) clonotypes were extracted and evaluated for their clonal abundance in phagemid DNA sets from round 0 (before bio-panning) to round 4. We then established a data set (TR data set) containing the antigen reactivity of scFv clones retrieved through TR technology and the clonal abundance of their HCDR3 and LCDR3 clonotypes in five sets of phagemid DNA. Using this TR data set, we trained our random forest (RF) machine learning algorithm to predict the binding properties of *in silico* HCDR3 and LCDR3 clonotypes [18,19].

To test the accuracy of our RF model (Figure 1), we extracted V_H and V_L sequences from MiSeq NGS data, which encompass both RF model-determined AR and antigen non-reactive (NR) HCDR3 or LCDR3 clonotypes and chemically synthesized them. Using these V_H and V_L genes,

we established two phage-displayed scFv libraries. The AR library was prepared using V_H and V_L genes encompassing AR HCDR3 and LCDR3 clonotypes, and the NR library was constructed using V_H and V_L genes encompassing NR HCDR3 and LCDR3 clonotypes. After one round of bio-panning on antigen-conjugated magnetic beads, antigen reactivity of phage clones was tested by phage ELISA. From the AR library, we obtained many scFv clones containing AR HCDR3 and LCDR3 clonotypes, while no AR clones were enriched from the NR library.

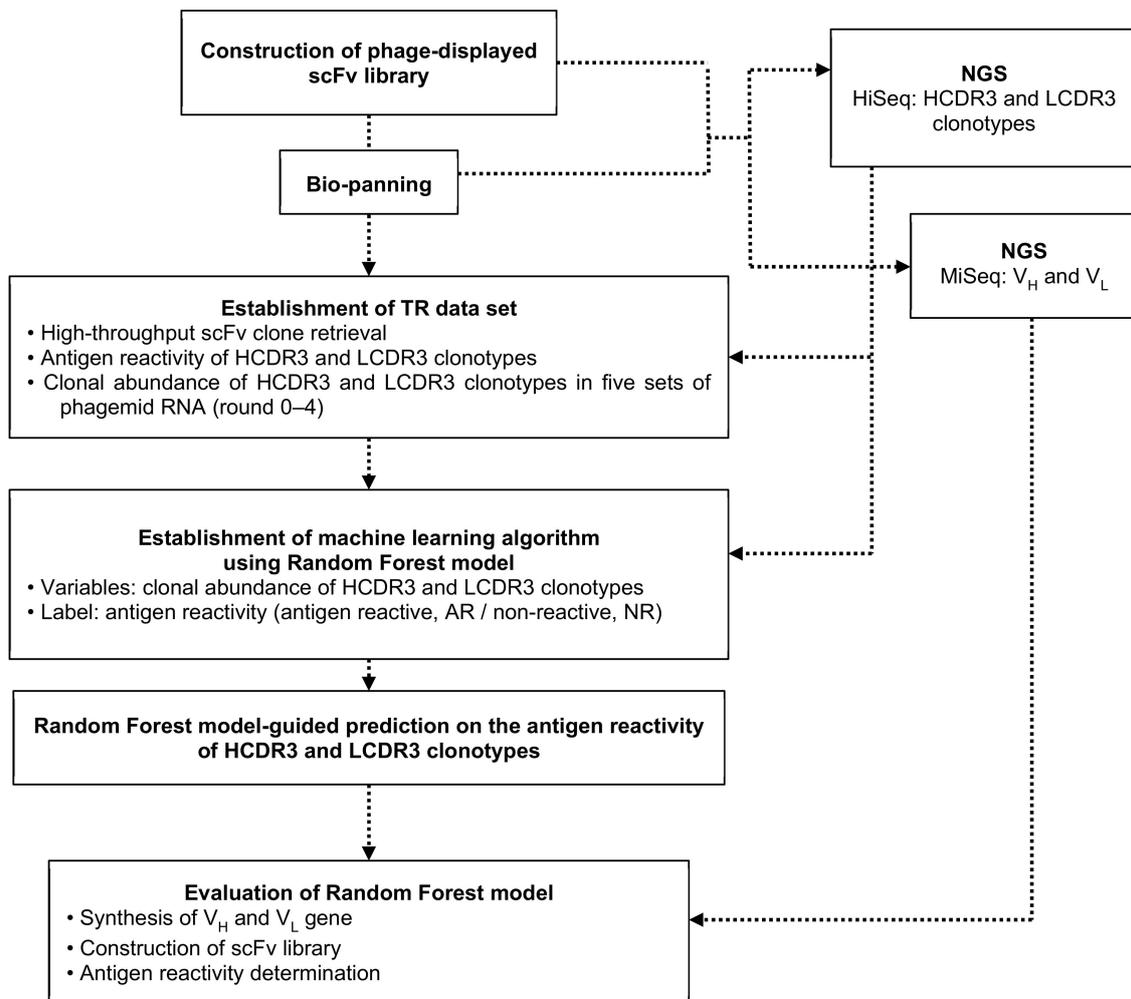


Figure 1. Workflow of the machine learning-guided selection of antigen-reactive HCDR3 and LCDR3 clonotypes with confirmation of their reactivity.

2. Materials and Methods

2.1. Immunization, Construction of Phage-Displayed scFv Library, and Bio-Panning

White leghorn chickens were immunized and boosted three times with 10 μ g of recombinant mouse c-Met-Fc chimera (527-ME; R&D systems, Carlsbad, CA, USA). The experiment was approved by the Ethics Committee of BioPOA, Ltd. (ethical approval code: BP-2019-C03-1). One week after the final boosting, total RNA was isolated from spleen, bone marrow, and bursa of Fabricius using TRIzol Reagent (15596018; Invitrogen), and cDNA was synthesized using SuperScript III first-strand cDNA synthesis kit with oligo dT priming (18418020; Invitrogen, Carlsbad, CA, USA). Using this cDNA, a phage-displayed scFv library was prepared as described previously [20,21]. V_H and V_L genes were amplified from the cDNA using specific primer sets utilized for the construction of scFv genes. Then, scFv genes were ligated into the pComb3XSS phage display vector, which was transfected into *E. coli* K12 ER2738 cells. Phage-displayed scFv libraries

were rescued from transfected cells after infection with VCSM13 helper phage and overnight culture, and then subjected to four rounds of bio-panning using recombinant mouse c-Met (50622-M08H, Sino Biological, Beijing, China)-conjugated magnetic beads (Dynabeads 14302D; Invitrogen). Antigen-coated magnetic beads were washed with 0.05% tween in phosphate-buffered saline (PBS) once for the first round, three times for the second and third rounds, and five times for the fourth round. After each round of bio-panning, phagemid DNA was prepared from bacterial cell pellets using a Qiaprep Spin Miniprep Kit (27104, Qiagen, Hilden, Germany).

2.2. Next-Generation Sequencing (NGS)

From five sets of phagemid DNA, short V_H and V_L gene fragments encoding the 3' part of FR3 and CDR3, and the 5' part of FR4, were amplified using primers designed to hybridize to FR3 and FR4 of the chicken V_H gene (LFR3: 5'-CCCTTCACGATTCTCCGGTGCC-3'; LFR4: 5'-CTGACCTAGGACGGT CAGGG-3'; HFR3: 5'-GGCTGCAGCTGAACAACCTCAGGGCTG-3'; HFR4: 5'-GGAGGAGACGA TGA CTTCGGTCCCGTGG-3'). Other gene fragments encoding the whole V_H and V_L genes were also amplified using specific primers previously described [16]. Prior to NGS analysis, all amplicon libraries were submitted for a quality control procedure on TapeStation 2200 (Agilent Technologies, Santa Clara, CA, USA). Libraries having a single peak of correct fragment length were subjected to NGS analysis using the HiSeq 2500 and MiSeq platforms (Illumina, Inc.) for short and whole V_H and V_L gene fragments, respectively. We uploaded the sequence data to NCBI (SRA accession number: PRJNA607865).

To ensure the quality of NGS data, the following pre-processing steps were performed. First, all pair-end reads were merged with PEAR using the developer's default parameters [22]. Second, we filtered out any reads that were compatible with the following description: (1) reads not meeting our minimum quality Phred score, (2) reads not having the primer sequence used in the phage-displayed scFv library construction process, (3) out-of-frame reads, and (4) reads without any identifiable CDR3. The reads were then collated based on their CDR3 sequences and any CDR3 clonotype with read count of less than 2 was discarded.

2.3. High-Throughput Clone Retrieval and Phage ELISA

The phagemid library from the final bio-panning round was transfected into *E. coli* K12 ER2738 cells, and then subjected to our high-throughput clonal retrieval procedure using TrueRepertoire (TR) technology, as described previously [17].

The retrieved phage clones were subjected to phage ELISA, as described previously with adequate modifications [21]. Phage clones were rescued overnight from the plate and culture supernatants containing phage that were diluted with equal volumes of 6% bovine serum albumin (BSA) solution in PBS. Phage solutions were then added to microtiter wells (3690, Corning life sciences, Corning, New York, NY, USA) coated with recombinant mouse c-Met or mouse anti-HA antibody (H3663, Merck, Darmstadt, Germany) and blocked with BSA. Microtiter plates were incubated for 2 h at 37 °C and washed three times with 0.05% Tween in PBS, which is followed by 3% BSA in PBS containing horseradish peroxidase (HRP)-conjugated anti-M13 antibody (11973-MM05, Sino Biological) in addition to each well. After incubation and washing as described above, HRP substrate solution 2,2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS) (002024, Thermo Fisher Scientific, Waltham, MA, USA) was added to each well. The plate was incubated 15 min and the absorbance values of each well were measured by a SkanIt microplate reader (Thermo Fisher Scientific) with a fast measurement protocol at a wavelength of 405 nm.

For each clone, the ratio (Relative Absorbance A) of the average absorbance of a recombinant mouse c-Met-coated well vs. an anti-HA antibody-coated well was calculated. The absorbance of an anti-HA-coated well was used to accommodate variations in the amount of phage in each phage clone. We also determined the ratio (Relative Absorbance B) of the average absorbance of a BSA-blocked well to an anti-HA antibody-coated well. When Relative Absorbance A exceeded +3 standard deviation of Relative Absorbance B, we designated the phage clone as antigen-reactive.

2.4. Establishment of the Random forest (RF) Models

Random forest (RF), regularized discriminant analysis (RDA), linear discriminant analysis (LDA), support vector machine (SVM), naïve bayes (NB), and AdaBoost (ADA) classification trees were selected for comparison. Our input data for the training of binder prediction models were created using a TR data set consisting of antigen reactivity for scFv clones found in TR analysis and the clonal abundance of their HCDR3 and LCDR3 clonotypes in five sets of phagemid DNA. The caret package for R was used to benchmark popular classification algorithms by their accuracy and Cohen's kappa value. Each algorithm was evaluated across five repetitions of 10-fold cross-validations (50 models in total). This meant that, for each repetition, the training data set was randomly divided into 10 parts and each of the 10 models were cross-validated by one unique part after being trained on the other nine parts of the training data set. No manual tuning was performed during this benchmarking phase [23].

To generate binder prediction models for HCDR3 and LCDR3 clonotypes using a random Forest package, we sampled a proportion of the TR data set without replacement to be used as a training data set for the RF model. The remaining portion of the TR data set served as a validation set to measure the performance of RF models [24]. The following parameters were adjusted to best tune our model's performance: (1) sampling ratio of training data, (2) number of variables (mtry) to randomly sample at each node of the decision-making tree, and (3) number of trees (ntree) to compromise our RF model. We then iterated through all combinations of parameters. Each combination was used to generate 10 different RF models to minimize any biases arising from the training data set not being representative of the TR data set. The validation set was then used to measure the performance of each RF model to determine optimal parameters for the RF. Using the randomForestExplainer package, the minimum depth of each variable was calculated, which is frequently used as a measure of variable importance to elucidate the decision-making process of the algorithm [25]. The minimum depth of a variable is defined as the distance between the root node and the variable's first appearance at a node of the decision tree. Thus, the variable with the smallest mean minimum depth could be regarded as the most important variable. To compare the variable importance results of the prediction model with actual experimental data, we tracked the enrichment pattern by measuring the bio-panning titer and clonal diversity change as Shannon's entropy (SE) [26] by following each round of bio-panning, as described previously [21].

2.5. Construction of Antigen-Reactive (AR) and Non-Reactive (NR) Phage-Displayed scFv Library and Phage ELISA

Forty antigen-reactive (AR) and 10 non-reactive (NR) V_H and V_L genes were chemically synthesized (Twist Bioscience, San Francisco, CA, USA). Forty AR V_H and 40 AR V_L genes were subjected to linker PCR to generate scFv genes, which were used to create the AR phage-displayed scFv library, as described previously [20]. In a parallel experiment, the NR phage-displayed scFv library was constructed using 10 NR V_H and 10 NR V_L genes. After a round of bio-panning using recombinant mouse c-Met (50622-M08H; Sino Biological)-conjugated magnetic beads (Dynabeads 14302D; Invitrogen) and washing once with 0.05% tween in PBS, 96 phage clones were randomly rescued from each AR and NR library and subjected to phage ELISA, as described above. After phage ELISA, the nucleotide sequences of scFv clones were determined by Sanger nucleotide sequencing (Macrogen, Seoul, South Korea).

3. Results

3.1. Construction of Phage-Displayed scFv Library, Biopanning, Selection of Positive Clones, Next-Generation Sequencing (NGS), And Establishment of TR Data Set

Chickens were immunized with recombinant mouse c-Met-Fc chimera. Spleen, bone marrow, and bursa of Fabricius were harvested from the immunized chickens and total RNA was prepared to generate a phage-displayed scFv library with a complexity of 4.96×10^9 . Four rounds of bio-panning were performed using antigen-coated magnetic beads. After the final round of bio-panning, the phage

pool was subjected to high-throughput clonal retrieval using TR technology. From the TR analysis, 641 clones with unique V_H and V_L pairs were identified. These phage clones were rescued and subjected to phage ELISA. Out of 641 phage clones, 149 clones showed reactivity to c-Met with statistical differences from non-reactive clones (data not shown) designated as AR clones. We used the binding reactivity of the 641 clones as a part of the TR data set.

After arranging the phage-displayed scFv library and each round of bio-panning, phagemid DNA (rounds 0–4) was prepared using bacterial pellets obtained after centrifugation of overnight culture supernatant. From these five sets of phagemid DNA, gene fragments encoding HCDR3 and LCDR3 were amplified and subjected to NGS analysis using the HiSeq platform. After NGS data pre-processing, we defined valid clonotypes as unique CDR3 sequences with read counts of two or higher in any set of phagemid DNA, and we were able to retrieve 860,207 HCDR3 clonotypes and 443,292 LCDR3 clonotypes across the entire bio-panning phase (Table 1). Clonal abundance throughout bio-panning stages was determined by counting the number of times that a clonotype appeared in each bio-panning round. The clonal abundance of clonotypes matching to scFv clones found in TR analysis was used as another part of the TR data set. We also amplified V_H and V_L gene fragments from five sets of phagemid DNA and subjected them to NGS analysis using the MiSeq platform.

Table 1. Number of CDR3 clonotypes obtained from the bio-panning procedure.

Clonotypes	Round 0	Round 1	Round 2	Round 3	Round 4	Total
HCDR3	390,814	395,459	402,854	311,678	308,547	860,207
LCDR3	272,317	253,899	250,630	187,314	117,239	443,292

3.2. Establishing Random Forest (RF) Binding Reactivity Prediction Model

We compared random forest, regularized discriminant analysis, linear discriminant analysis, support vector machine, naïve bayes, and AdaBoost classification trees for their accuracy and kappa score distributions. We found out that the random forest algorithm was best suited for binder predictions of HCDR3 clonotypes with the mean accuracy of 89.69% and mean Cohen’s kappa value of 0.45 (Tables S1–S4 and Figure S1). While regularized discriminant analysis did perform marginally better in the LCDR3 clonotypes, random forest showed more potential for improvement with manual tuning when consulting maximum accuracy and Cohen’s kappa value. With these observations, we decided to adopt random forest models to establish a binding reactivity prediction model.

Utilizing the TR data set, two separate RF models were trained for HCDR3 and LCDR3 clonotypes. The algorithm was instructed to treat the clonal abundance of clonotypes in the five sets of phagemid DNA (round 0–4) as predictor variables and the binding reactivity as the response variable. Thus, each unique clonotype in our TR data set was individually labelled with that clonotype’s abundance at each of the bio-panning rounds and its binding reactivity. Before the training of each new RF model, the TR data set was divided into a training data set and a validation data set. After training the RF model using the training set, the validation set was presented to the RF model, and RF model accuracy in predicting clonotype binding reactivity was determined.

To determine the optimum training parameters for our RF model, 7200 RF models were evaluated. Optimizing for sensitivity, the ideal parameters for the HCDR3 RF model were found to be a 75% sampling ratio of the TR data set, mtry of 4, and ntree of 500. The performance metrics of 10 RF models using those parameters were: (1) mean accuracy of 90.48%, (2) mean sensitivity of 44.36%, and (3) mean specificity of 97.61%. Optimizing for accuracy, the ideal parameters for the LCDR3 RF model were found to be a 65% sampling ratio of the TR data set, mtry of 2, and ntree of 500. Once again, the performance metrics of 10 LCDR3 RF models using those parameters were: (1) mean accuracy of 86.47%, (2) sensitivity of 55.98%, and (3) specificity of 94.90% (Table S5).

3.3. Measurement of the Minimum Depth Value of a Predictor Variable

The minimum depth of a predictor variable can be interpreted as a measure of the variable importance. We extracted the minimum depth value of each predictor variable from the 500 decision trees that compromised our RF model. Of note is that, in our HCDR3 RF model, our predictor variable representing CDR3 clonal abundance in round 3 of bio-panning was most likely to appear at the root node of our decision trees appearing in 360 instances out of our 500 decision trees, and, consequently, had the lowest mean minimum depth of 0.46. In our LCDR3 RF model, our predictor variable representing CDR3 clonal abundance in round 4 of bio-panning was most likely to appear at the root node of our decision trees appearing in 195 instances out of 500 decision trees and, consequently, had the lowest minimum depth of 1.16 (Figure 2). In accordance with these observations, the Shannon entropy (SE) representing clonal diversity dropped at round 3 in the case of HCDR3 clonotypes while the SE of LCDR3 significantly dropped at round 4 (Table S6 and Figure S2).

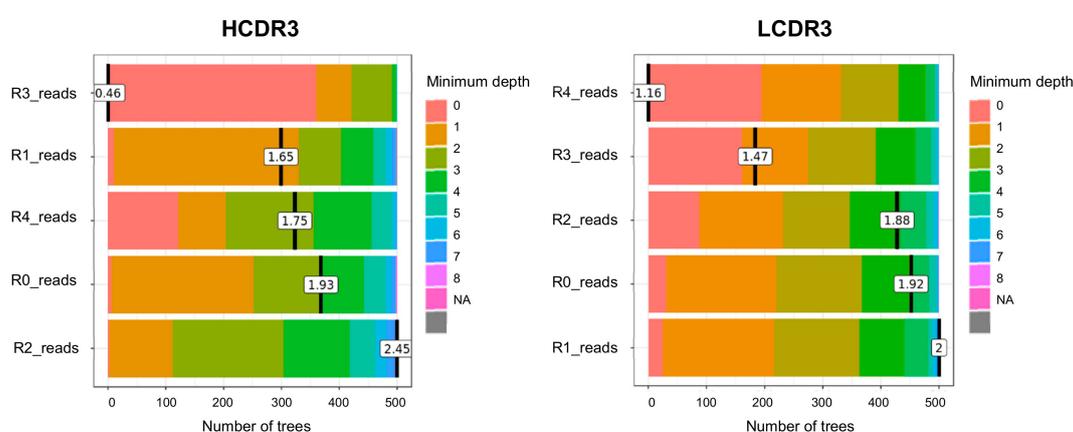


Figure 2. Distribution of the minimum depth of predictor variables (clonal abundance at round 0–4 of bio-panning) from individual decision trees in the RF prediction model for CDR3 clonotypes. Minimum depth value is colored according to its depth and mean value is calculated and displayed at points.

We also observed the interaction of our predictor variables taking place within the decision trees. Variable interactions are regarded as taking a sub-tree of two nodes and considering it as a single node. We can then look at the minimum depth value of that sub-tree to gauge the interaction's importance in classifying its input. In our HCDR3 RF model, the top four most influential interactions all involved clonal abundance in round 3 of bio-panning as the root node. The most influential interaction took place between round 3 and round 1 with a minimum depth value of 0.84 (Table S7). In our LCDR3 RF model, three of the top four most influential interactions involved clonal abundance in round 4 of bio-panning as the root node. The most influential interaction took place between round 4 and round 0 with a minimum depth value of 1.18 (Table S8). Using the training data set, the clonal abundance of HCDR3 clonotype in round 3 and round 1 and that of LCDR3 clonotype in round 4 and round 0 were plotted in Figure 3a,b, which shows significant correlation.

3.4. Predicting the Binding Property of the CDR3 Clonotype Using RF Modeling

Of the 860,207 HCDR3 clonotypes fed into the RF model, 5,780 clonotypes were predicted to be AR. Of the 443,292 LCDR3 clonotypes, 34,703 clonotypes were predicted to be AR. The confidence value of the RF model for each prediction was also obtained. For HCDR3 and LCDR3, 1.70% (98/5,780) and 0.16% (58/34,703) of clonotypes, respectively, were predicted to be AR with a confidence value of more than 0.9. Meanwhile, 0.56% (4,825/854,427) of HCDR3 clonotypes and 41.14% (168,116/408,589) of LCDR3 clonotypes were predicted to be NR with a confidence value over 0.9. When CDR3 clonotypes were visualized with the most important variable interactions together including a confidence value

(Figure 3c), clonotypes with higher confidence values were distributed near the axis of the most important variable akin to the distribution of AR clonotypes in the training data set.

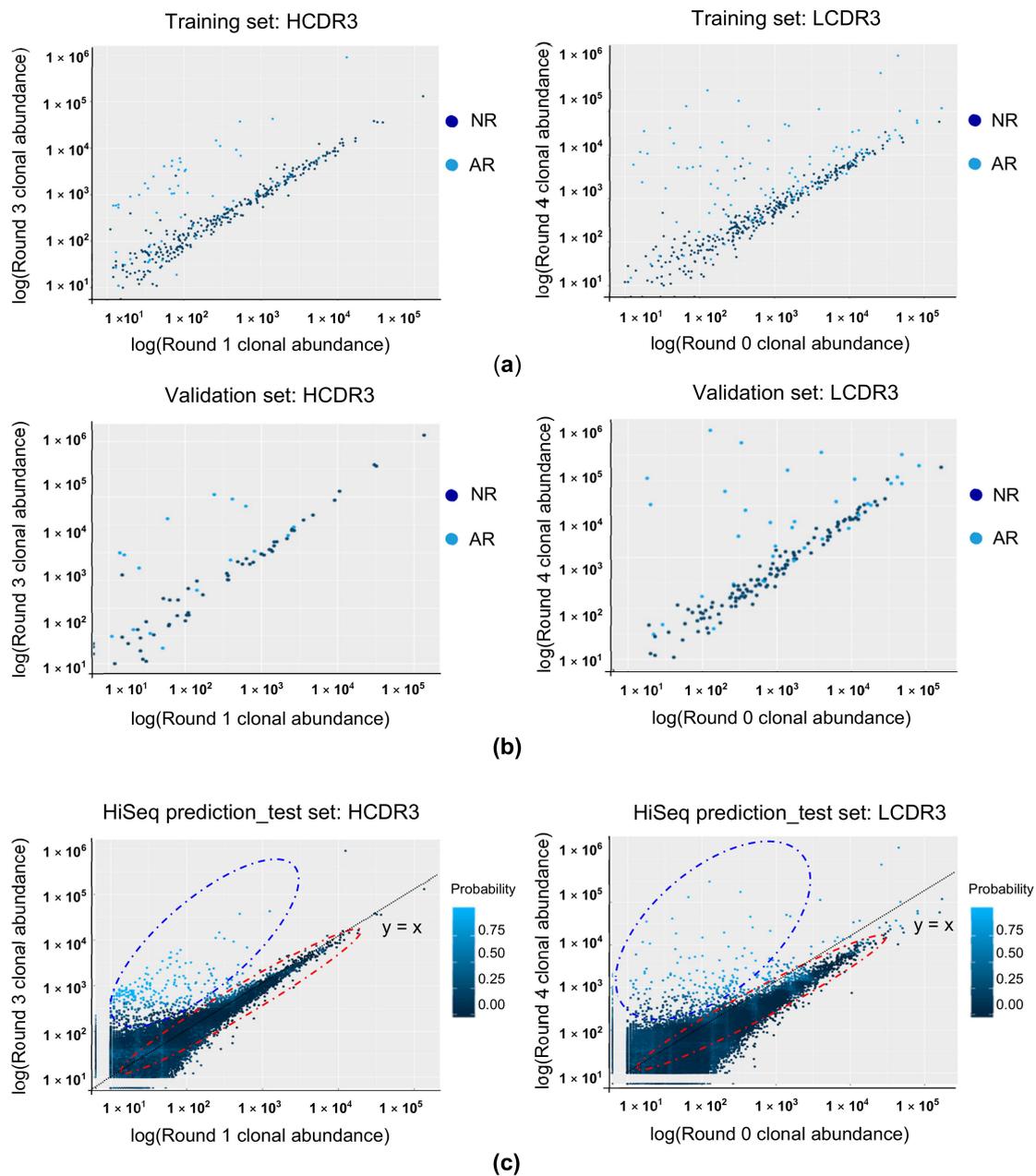


Figure 3. The most influential variable interactions and distributions of CDR3 clonotypes. (a) Clonal abundance at the most influential interaction is plotted with binding property label from training data used in the random forest (RF) prediction model. AR, antigen-reactive, NR, antigen non-reactive. (b) Clonal abundance at the most influential interaction is plotted with a binding property label from validation data used in the RF prediction model. (c) Clonal abundance at the most influential interaction is plotted with confidence value (probability) from HiSeq-identified CDR3 clonotypes. Clonotypes with higher confidence values are distributed near the root variable axis (highlighted with a dashed blue circle) while clonotypes having lower confidence values are distributed below the $y = x$ axis (dotted line) (highlighted with a dashed red circle).

3.5. Antigen Reactivity Validation of *In Silico* CDR3 Clonotypes in Phage ELISA

We selected 40 HCDR3 AR, 40 LCDR3 AR, 10 HCDR3 NR, and 10 LCDR3 NR clonotypes with the highest confidence values of which whole V_H or V_L gene sequences were available from the NGS data generated from five sets of phagemid DNA using the MiSeq platform (Tables S9 and S10). After whole V_H and V_L genes were chemically synthesized, V_H and V_L genes of AR clonotypes were used to construct the AR phage-displayed scFv library. In a parallel experiment, the NR phage-displayed scFv library was also constructed using the same scheme. After a single round of bio-panning on antigen-coated magnetic beads, 96 phage clones were randomly selected from the output titer plate of the AR library and subjected to phage ELISA. Fifteen phage clones were found to be AR, which turned out to be 14 scFv clones consisting of five HCDR3 and 11 LCDR3 clonotypes by Sanger sequencing (Figure 4, Table 2). AR5 and AR6 phage clones encoded the same scFv sequence. It was noticeable that three LCDR3 clonotypes were paired with two different HCDR3 clonotypes as in AR1 and AR13, AR2, and AR7, and AR4 and AR14 phage clones showing light chain redundancy. In a parallel experiment, no AR clones were identified from 96 phage clones from the NR library. Sixteen clones were randomly selected and Sanger sequencing was performed to find 13 HCDR3 and nine LCDR3 clonotypes. With these results, we concluded that our RF model can be used to select HCDR3 and LCDR3 AR clonotypes with a significant hit ratio.

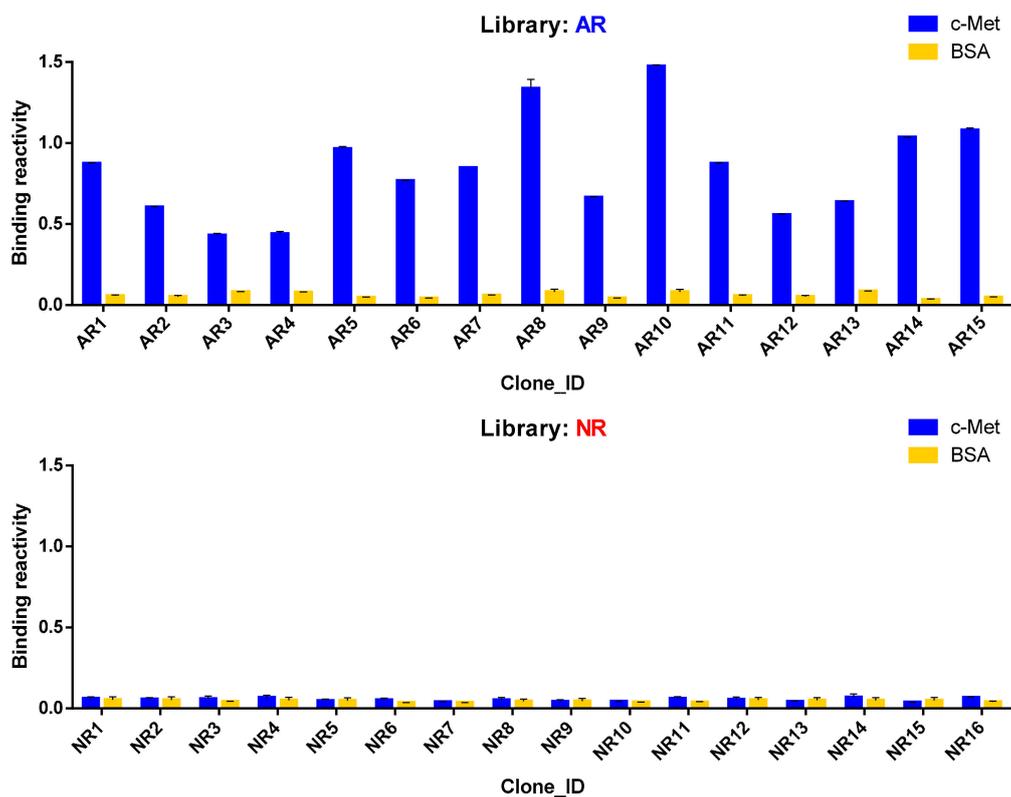


Figure 4. Reactivity of phage-displayed scFv clones in phage ELISA. Binding reactivity of 15 unique clones identified from the AR library and 16 unique clones from the NR library are shown. Wells in microtiter plates were either coated with recombinant mouse c-Met or just blocked with 3% BSA in PBS. Phage clones, HRP-conjugated anti-M13 antibody, and HRP substrate solution were added sequentially with intermittent washing.

Table 2. Amino acid sequences of AR CDR3 clonotypes identified from AR library.

Clone ID	HCDR3 AA * Sequence	LCDR3 AA * Sequence
AR1	GSGGVDSIDA	GSYDNTYAGI
AR2	SADGYGWDTAGNMDA	GSIDSNYDGI
AR3	TAGTCTTSCNAGAYIDA	GGYDSSAA
AR4	TTCSGSYGWCADSIDA	GAYDSSYIGI
AR5	SADSCATCATYPSEIDT	GSFDSSYVGM
AR6	SADSCATCATYPSEIDT	GSFDSSYVGM
AR7	SADSCATCATYPSEIDT	GSIDSNYDGI
AR8	SADSCATCATYPSEIDT	GSYDSSYVGL
AR9	SADSCATCATYPSEIDT	GSYDSSYDGV
AR10	SADSCATCATYPSEIDT	GSFDSSYTGI
AR11	SADSCATCATYPSEIDT	GSIDSRVGI
AR12	SADSCATCATYPSEIDT	GSYDSSYVGVGV
AR13	SADSCATCATYPSEIDT	GSYDNTYAGI
AR14	SADSCATCATYPSEIDT	GGYDSSGA
AR15	SADSCATCATYPSEIDT	GAYDSSYIGI

* AA: amino acid.

4. Discussion

Despite the promise of targeting the HGF-c-Met signaling pathway for cancer therapy, no specific therapeutic agent has been approved for clinical use. Small molecule inhibitors specific to c-Met are yet to be approved, and only nonspecific tyrosine kinase inhibitors inhibiting c-Met are available (Table 3) [27]. Recombinant protein (truncated HGF, decoy c-Met) was not successful in clinical trials due to several factors, including short half-life and low target affinity limiting the intended efficacy [28]. Several HGF-neutralizing antibodies have been developed with two currently active in clinical trials [29]. However, the inhibitory targeting of c-Met by an antibody has been difficult since the bivalency of antibodies often induces receptor dimerization, which potentially causes cancer cell proliferation and migration. As such, both a monovalent form of antibody blocking its interaction with HGF and a bivalent antibody inducing receptor internalization have been developed and tested in clinical trials unsuccessfully [13,30]. Recently, an anti-EGFR × c-Met bispecific antibody monovalent to each target came under clinical development, which should inhibit the ligand interaction and induce the internalization of both receptors [31,32]. Besides blocking the interaction with ligand and receptor internalization, other mechanisms of actions for therapeutic antibody binding to targets on cancer cells were also reported, which include complement-dependent cell cytotoxicity as observed in rituximab [33], antibody-dependent cell cytotoxicity seen with obinutuzumab [34], and phagocytosis of antibody-opsonized tumor cells [35]. Antibodies are also used to deliver cytotoxic payloads into cancer cells such as with T-DM1 [36], and cross-linking cancer cells to cytotoxic T cells with blinatumomab [37]. Furthermore, antibodies are used as a cancer cell-targeting component in chimeric antigen receptor T cell therapy, as seen with tisagenlecleucel and axicabtagene ciloleucel [38]. Additionally, it is well known that the antibody epitope and binding characteristics critically influence efficacy for all these various modes of action [39,40]. Therefore, it is crucial to develop a significant number of antibodies to a target and characterize their performance. However, antibody selection technologies, including conventional hybridoma and display technologies such as phage, ribosomal, and bacterial, all have their own limitations regarding high-throughput capabilities.

After George P. Smith and Gregory P. Winter successfully displayed recombinant peptides and antibodies at the pIII protein of the M13 phage [41], this powerful technology has evolved and been actively applied toward therapeutic antibody discovery [42,43]. Currently, over 80 antibodies derived from phage display libraries have entered clinical studies with 10 of these granted marketing authorization [44]. Since Ravn U et al. demonstrated the potential for NGS analysis in the phage-displayed antibody repertoire in 2010, numerous groups have leveraged similar strategies for discovering antibodies reactive to specific antigens [16,45–53]. The next hurdle to overcome after the

identification of in silico antibody sequences in NGS data was the low-throughput nature of chemically synthesizing all antibody sequences and individually testing their reactivity. Recently, we introduced a method for combining NGS analysis and individual antibody sequence identification with the isolation of their physical DNA, which was named TR technology [17]. To reduce the burden of expressing all of the antibodies, we also devised a way of predicting antigen reactivity toward antigens by clustering antibody clonotypes with their patterns of enrichment or restriction through bio-panning rounds, and then combining TR with clustering and testing reactivity for a significant number of clones.

Table 3. Clinical usage of small molecule inhibitors targeting c-Met in cancer therapy.

Drug Name	Targets	FDA Approval Status	Approved Year
Tivantinib	c-Met, microtubule	None	N.A.*
Foretinib	c-Met, VEGFR-2 *	None	N.A.
Cabozantinib	c-Met, VEGFR, Axl	Medullary thyroid cancer	2012
		Advanced renal cell carcinoma	2016
Crizotinib	c-Met, ALK *, ROS1, RON *	Hepatocellular carcinoma	2019
Capmatinib	c-Met, EGFR *, ErbB-3	ALK or ROS-1 positive NSCLC *	2011
AMG337	c-Met	None	N.A.
AZD6094	c-Met	None	N.A.
BMS777607/ASLAN002	c-Met, Axl, Tyro3, RON	None	N.A.
Glesatinib	c-Met, Axl	None	N.A.
Tepotinib	c-Met	None	N.A.

* VEGFR-2: Vascular endothelial growth factor-2, ALK: Anaplastic lymphoma kinase, RON: Receptor d'Origine nantais, EGFR: Epidermal growth factor receptor, NSCLC: Non-small cell lung cancer, N.A.: not available.

Using these tools and procedures, we believed that it was possible to train a machine learning algorithm to derive in silico AR clonotypes from a repertoire of NGS sequences. To demonstrate this, we performed an in-depth analysis of our bio-panning library with the guidance of our supervised machine learning algorithm trained with large amounts of data sets generated from a high-throughput clone retrieval platform and independent NGS analysis. The RF model utilized is composed of numerous unique decision trees that work together to classify inputs. Each decision tree in an RF model is generated using a bootstrapped sample of the training data and a randomized subset of variables evaluated for the best split at each node of that decision tree. As a result, each RF model decision tree is uniquely generated and makes the model more robust to overfitting compared to other linear classifiers or decision trees. Compared to the more complicated black boxes of artificial neural networks, RF models frequently show similar levels of predictive performance while remaining observable and transparent. By inspecting the composition of decision trees in the RF model, we can extract important measures of input variables to better understand the decision-making process of the algorithm. Our extraction of variable importance measures helped explain the logical processes of our RF prediction model, which consists of complex, randomized interactions of predictor variables and response variables. From these results, we can infer that AR HCDRs are mostly selected in enrichment rounds, while LCDR3s are significantly enriched with selected HCDR3s after additional selective pressure occurs. We can then infer that enrichment of scFv molecules depends on individual chains in different stages of the bio-panning process (first V_H is then significantly biased by V_L). We believe our prediction model may be enhanced to better predict binding reactivity with multiple (high, mid, low) rather than binary (reactive/non-reactive) classifications. It is highly likely that this model can be applied to other display platforms that use bio-panning as the selection process, such as yeast display library for fluorescence-activated cell sorting screening [54]. Recently, artificial intelligence has been applied to predict the physicochemical properties of antibody sequences [55–59] and/or optimize them [60–62].

In summary, we report that machine learning algorithm can provide a way to identify AR antibody clones with a significant hit ratio, which will allow us to better characterize diverse antibodies in greater numbers currently unattainable by traditional methods.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2218-273X/10/3/421/s1>, Figure S1: Evaluation of 6 prediction models using training data sets, Figure S2: Shannon's entropy (SE) change following biopanning procedure, Table S1: Accuracy score distributions of random forest (RF), regularized discriminant analysis (RDA), linear discriminant analysis (LDA), support vector machines (SVM), naïve bayes (NB), AdaBoost Classification Trees (ADA) for HCDR3 binding reactivity predictions, Table S2: Kappa score distributions of RF, RDA, LDA, SVM, NB, and ADA for HCDR3 binding reactivity predictions, Table S3: Accuracy score distributions of RF, RDA, LDA, SVM, NB, and ADA for LCDR3 binding reactivity predictions, Table S4: Kappa score distributions of RF, RDA, LDA, SVM, NB, and ADA for LCDR3 binding reactivity predictions, Table S5: Optimal parameter tuning in generation of random forest model, Table S6: Biopanning titer following four rounds of biopanning, Table S7: Mean-minimal depth of each variables and interaction (HCDR3), Table S8: Mean-minimal depth of each variables and interaction (LCDR3), Table S9: Predicted clones with HCDR3, full variable domain sequences with prediction results and confidence value, Table S10: Predicted clones with LCDR3, full variable domain sequences with prediction results and confidence value.

Author Contributions: D.K.Y. and J.C. (Junho Chung). Experiment, D.K.Y. and H.K.L. Methodology, D.K.Y. and J.C. (Junho Chung). Software analysis, S.R.L., H.H., and Y.J. Validation, D.K.Y. and S.R.L. Data curation, D.K.Y., S.R.L., and J.C. (Junho Chung). Visualization, D.K.Y. and S.R.L. Writing – original draft preparation, D.K.Y. and S.R.L. Writing – review and editing, J.H., S.K., J.C. (Jisu Chae), Y.J., and T.R. Supervision, J.C. (Junho Chung) and T.R. Project administration, J.C. (Junho Chung). Funding acquisition, J.C. (Junho Chung). All authors have read and agreed to the published version of the manuscript.

Funding: The Korea Drug Development Fund, Korea (Grant Number: KDDF-201904-19), supported this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Comoglio, P.; Trusolino, L.; Boccaccio, C. Known and novel roles of the MET oncogene in cancer: a coherent approach to targeted therapy. *Nat. Rev. Cancer* **2018**, *18*, 341–358. [[CrossRef](#)] [[PubMed](#)]
2. Lai, G.G.; Lim, T.H.; Lim, J.; Liew, P.J.; Kwang, X.L.; Nahar, R.; Aung, Z.W.; Takano, A.; Lee, Y.Y.; Lau, D.P.; et al. Clonal MET Amplification as a Determinant of Tyrosine Kinase Inhibitor Resistance in Epidermal Growth Factor Receptor-Mutant Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **2019**, *37*, 876–884. [[CrossRef](#)] [[PubMed](#)]
3. Park, J.; Choi, Y.; Ko, Y.S.; Kim, Y.; Pyo, J.-S.; Jang, B.G.; A Kim, M.; Lee, J.-S.; Chang, M.S.; Park, J.-W.; et al. FOXO1 Suppression is a Determinant of Acquired Lapatinib-Resistance in HER2-Positive Gastric Cancer Cells Through MET Upregulation. *Cancer Res. Treat.* **2018**, *50*, 239–254. [[CrossRef](#)] [[PubMed](#)]
4. Pietrantonio, F.; Fucà, G.; Morano, F.; Gloghini, A.; Corso, S.; Aprile, G.; Perrone, F.; De Vita, F.; Tamborini, E.; Tomasello, G.; et al. Biomarkers of Primary Resistance to Trastuzumab in HER2-Positive Metastatic Gastric Cancer Patients: the AMNESIA Case-Control Study. *Clin. Cancer Res.* **2017**, *24*, 1082–1089. [[CrossRef](#)] [[PubMed](#)]
5. Dagogo-Jack, I.; Shaw, A.T. Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* **2017**, *15*, 81–94. [[CrossRef](#)] [[PubMed](#)]
6. Martin, V.; Chiriaco, C.; Modica, C.; Acquadro, A.; Cortese, M.; Galimi, F.; Perera, T.; Gammaitoni, L.; Aglietta, M.; Comoglio, P.M.; et al. Met inhibition revokes IFN γ -induction of PD-1 ligands in MET-amplified tumours. *Br. J. Cancer* **2019**, *120*, 527–536. [[CrossRef](#)]
7. Saigí, M.; Albuquerque-Bejar, J.J.; Mc Leer-Florin, A.; Pereira, C.; Pros, E.; Romero, O.A.; Baixeras, N.; Esteve-Codina, A.; Nadal, E.; Brambilla, E.; et al. MET-Oncogenic and JAK2-Inactivating Alterations Are Independent Factors That Affect Regulation of PD-L1 Expression in Lung Cancer. *Clin. Cancer Res.* **2018**, *24*, 4579–4587. [[CrossRef](#)]
8. Glodde, N.; Bald, T.; Boorn-Konijnenberg, D.V.D.; Nakamura, K.; O'Donnell, J.; Szczepanski, S.; Brandes, M.; Eickhoff, S.; Das, I.; Shridhar, N.; et al. Reactive Neutrophil Responses Dependent on the Receptor Tyrosine Kinase c-MET Limit Cancer Immunotherapy. *Immunity* **2017**, *47*, 789–802.e9. [[CrossRef](#)]
9. Papaccio, F.; Della Corte, C.M.; Viscardi, G.; Di Liello, R.; Esposito, G.; Sparano, F.; Ciardiello, F.; Morgillo, F. HGF/MET and the Immune System: Relevance for Cancer Immunotherapy. *Int. J. Mol. Sci.* **2018**, *19*, 3595. [[CrossRef](#)]

10. Lee, J.; Kim, S.; Nam, D.-H.; Lee, S.J.; Park, S.H.; Park, J.O.; Lee, J.-W.; Kim, K.-M.; Lee, H.; Park, N.; et al. First-in-human phase I trial of anti-hepatocyte growth factor (HGF) antibody (YYB101) in refractory solid tumor patients: Integrative pathologic-genomic analysis and the final results. *J. Clin. Oncol.* **2019**, *37*, 3104. [[CrossRef](#)]
11. Rolfo, C.; Van Der Steen, N.; Pauwels, P.; Cappuzzo, F. Onartuzumab in lung cancer: the fall of Icarus? *Expert Rev. Anticancer. Ther.* **2015**, *15*, 487–489. [[CrossRef](#)] [[PubMed](#)]
12. Parola, C.; Neumeier, D.; Reddy, S. Integrating high-throughput screening and sequencing for monoclonal antibody discovery and engineering. *Immunology* **2017**, *153*, 31–41. [[CrossRef](#)] [[PubMed](#)]
13. Brown, A.; Snapkov, I.; Akbar, R.; Pavlović, M.; Miho, E.; Sandve, G.K.; Greiff, V. Augmenting adaptive immunity: progress and challenges in the quantitative engineering and analysis of adaptive immune receptor repertoires. *Mol. Syst. Des. Eng.* **2019**, *4*, 701–736. [[CrossRef](#)]
14. A Norman, R.; Ambrosetti, F.; Bonvin, A.M.J.J.; Colwell, L.J.; Kelm, S.; Kumar, S.; Krawczyk, K. Computational approaches to therapeutic antibody design: established methods and emerging trends. *Briefings Bioinform.* **2019**. [[CrossRef](#)] [[PubMed](#)]
15. Rouet, R.; Jackson, K.J.L.; Langley, D.B.; Christ, D. Next-Generation Sequencing of Antibody Display Repertoires. *Front. Immunol.* **2018**, *9*, 9–118. [[CrossRef](#)]
16. Yang, W.; Yoon, A.; Lee, S.; Kim, S.; Han, J.; Chung, J. Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. *Exp. Mol. Med.* **2017**, *49*, e308. [[CrossRef](#)]
17. Noh, J.; Kim, O.; Jung, Y.; Han, H.; Kim, J.-E.; Kim, S.; Lee, S.; Park, J.; Jung, R.H.; Kim, S.I.; et al. High-throughput retrieval of physical DNA for NGS-identifiable clones in phage display library. *mAbs* **2019**, *11*, 532–545. [[CrossRef](#)]
18. Singh, A.; Thakur, N.; Sharma, A. A review of supervised machine learning algorithms. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 1310–1315.
19. Kulkarni, V.Y.; Sinha, P.K. Effective Learning and Classification using Random Forest Algorithm. *Int. J. Eng. Innov. Technolgy* **2014**, *3*, 267–273.
20. Andris-Widhopf, J.; Rader, C.; Steinberger, P.; Fuller, R.; Barbas, C.F. Methods for the generation of chicken monoclonal antibody fragments by phage display. *J. Immunol. Methods* **2000**, *242*, 159–181. [[CrossRef](#)]
21. Barbas, C.F., III; Burton, D.R.; Scott, J.K.; Silverman, G.J. *Phage Display: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Suffolk County, NY, USA, 2001.
22. Zhang, J.; Kobert, K.; Flouri, T.; Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **2013**, *30*, 614–620. [[CrossRef](#)]
23. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 5. [[CrossRef](#)]
24. Liaw, A.; Wiener, M. Classification and regression by randomforest. *R News* **2002**, *2*, 18–22.
25. Wright, M.N.; Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 77. [[CrossRef](#)]
26. Rempala, G.A.; Seweryn, M.; Seweryn, M. Methods for diversity and overlap analysis in T-cell receptor populations. *J. Math. Boil.* **2012**, *67*, 1339–1368. [[CrossRef](#)] [[PubMed](#)]
27. Hughes, V.S.; Siemann, D.W. Have Clinical Trials Properly Assessed c-Met Inhibitors? *Trends Cancer* **2018**, *4*, 94–97. [[CrossRef](#)] [[PubMed](#)]
28. Moosavi, F.; Giovannetti, E.; Saso, L.; Firuzi, O. HGF/MET pathway aberrations as diagnostic, prognostic, and predictive biomarkers in human cancers. *Crit. Rev. Clin. Lab. Sci.* **2019**, *56*, 533–566. [[CrossRef](#)]
29. Kim, K.H.; Kim, H. Progress of antibody-based inhibitors of the HGF-cMET axis in cancer therapy. *Exp. Mol. Med.* **2017**, *49*, e307. [[CrossRef](#)]
30. Oh, Y.M.; Song, Y.-J.; Lee, S.B.; Jeong, Y.; Kim, B.; Kim, G.W.; Kim, K.E.; Lee, J.M.; Cho, M.-Y.; Choi, J.; et al. A New Anti-c-Met Antibody Selected by a Mechanism-Based Dual-Screening Method: Therapeutic Potential in Cancer. *Mol. Cells* **2012**, *34*, 523–529. [[CrossRef](#)]
31. Patnaik, A.; Gordon, M.; Tsai, F.; Papadopoulos, K.P.; Rasco, D.; Beeram, M.; Fu, S.; Janku, S.; Hynes, S.M.; Gundala, S.R.; et al. A phase I study of LY3164530, a bispecific antibody targeting MET and EGFR, in patients with advanced or metastatic cancer. *Cancer Chemother. Pharmacol.* **2018**, *82*, 407–418. [[CrossRef](#)]

32. Park, K.; Ahn, M.; Lee, S.; Kim, H.; Hong, M.; Millington, D.; Curtis, M.; Triantos, S.; Chaplan, S.; Haddish-Berhane, N.; et al. OA10.06 A First-in-Human Phase 1 Trial of the EGFR-cMET Bispecific Antibody JNJ-61186372 in Patients with Advanced Non-Small Cell Lung Cancer (NSCLC). *J. Thorac. Oncol.* **2018**, *13*, S344–S345. [[CrossRef](#)]
33. Pierpont, T.M.; Limper, C.B.; Richards, K.L. Past, Present, and Future of Rituximab—The World’s First Oncology Monoclonal Antibody Therapy. *Front. Oncol.* **2018**, *8*, 8. [[CrossRef](#)]
34. Zahavi, D.; Aldeghaither, D.; O’Connell, A.; Weiner, L.M. Enhancing antibody-dependent cell-mediated cytotoxicity: a strategy for improving antibody-based immunotherapy. *Antib. Ther.* **2018**, *1*, 7–12. [[CrossRef](#)]
35. Saffi, G.T.; Botelho, R.J. Lysosome Fission: Planning for an Exit. *Trends Cell Boil.* **2019**, *29*, 635–646. [[CrossRef](#)]
36. Rinnerthaler, G.; Gampenrieder, S.P.; Greil, R. HER2 Directed Antibody-Drug-Conjugates beyond T-DM1 in Breast Cancer. *Int. J. Mol. Sci.* **2019**, *20*, 1115. [[CrossRef](#)]
37. Romero, D. Haematological cancer: Blinatumomab facilitates complete responses. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 200.
38. Rafiq, S.; Hackett, C.S.; Brentjens, R.J. Engineering strategies to overcome the current roadblocks in CAR T cell therapy. *Nat. Rev. Clin. Oncol.* **2019**, *17*, 147–167. [[CrossRef](#)]
39. Sela-Culang, I.; Kunik, V.; Ofran, Y. The Structural Basis of Antibody-Antigen Recognition. *Front. Immunol.* **2013**, *4*, 302. [[CrossRef](#)]
40. Jespersen, M.C.; Mahajan, S.; Peters, B.; Nielsen, M.; Marcatili, P. Antibody Specific B-Cell Epitope Predictions: Leveraging Information From Antibody-Antigen Protein Complexes. *Front. Immunol.* **2019**, *10*, 298. [[CrossRef](#)]
41. Smith, G.P.; Petrenko, V.A. Phage Display. *Chem. Rev.* **1997**, *97*, 391–410. [[CrossRef](#)]
42. Mimmi, S.; Maisano, D.; Quinto, I.; Iaccino, E. Phage Display: An Overview in Context to Drug Discovery. *Trends Pharmacol. Sci.* **2019**, *40*, 87–91. [[CrossRef](#)]
43. Peltomaa, R.; Benito-Pena, E.; Barderas, R.; Bondi, M.C.M. Phage Display in the Quest for New Selective Recognition Elements for Biosensors. *ACS Omega* **2019**, *4*, 11569–11580. [[CrossRef](#)]
44. Kaplon, H.; Muralidharan, M.; Schneider, Z.; Reichert, J.M. Antibodies to watch in 2020. *mAbs* **2019**, *12*, 219–238. [[CrossRef](#)]
45. Ravn, U.; Gueneau, F.; Baerlocher, L.; Osteras, M.; Desmurs, M.; Malinge, P.; Magistrelli, G.; Farinelli, L.; Kosco-Vilbois, M.H.; Fischer, N. By-passing in vitro screening—next generation sequencing technologies applied to antibody display and in silico candidate selection. *Nucleic Acids Res.* **2010**, *38*, e193. [[CrossRef](#)] [[PubMed](#)]
46. Ravn, U.; Didelot, G.; Venet, S.; Ng, K.-T.; Gueneau, F.; Rousseau, F.; Calloud, S.; Kosco-Vilbois, M.; Fischer, N. Deep sequencing of phage display libraries to support antibody discovery. *Methods* **2013**, *60*, 99–110. [[CrossRef](#)] [[PubMed](#)]
47. D’Angelo, S.; Kumar, S.; Naranjo, L.; Ferrara, F.; Kiss, C.; Bradbury, A. From deep sequencing to actual clones. *Protein Eng. Des. Sel.* **2014**, *27*, 301–307. [[CrossRef](#)] [[PubMed](#)]
48. Hu, N.; Hu, S.; Wan, W.; Xu, M.; Du, R.; Zhao, W.; Gao, X.; Liu, J.; Liu, H.; Hong, J. Effective Optimization of Antibody Affinity by Phage Display Integrated with High-Throughput DNA Synthesis and Sequencing Technologies. *PLOS ONE* **2015**, *10*, e0129125. [[CrossRef](#)]
49. Spiliotopoulos, A.; Owen, J.P.; Maddison, B.C.; Dreveny, I.; Rees, H.; Gough, K. Sensitive recovery of recombinant antibody clones after their in silico identification within NGS datasets. *J. Immunol. Methods* **2015**, *420*, 50–55. [[CrossRef](#)]
50. Miyazaki, N.; Kiyose, N.; Akazawa, Y.; Takashima, M.; Hagihara, Y.; Inoue, N.; Matsuda, T.; Ogawa, R.; Inoue, S.; Ito, Y.; et al. Isolation and characterization of antigen-specific alpaca (*Lama pacos*) VHH antibodies by biopanning followed by high-throughput sequencing. *J. Biochem.* **2015**, *158*, 205–215. [[CrossRef](#)]
51. Lövgren, J.; Pursiheim, J.-P.; Pyykkö, M.; Salmi, J.; Lamminmäki, U. Next generation sequencing of all variable loops of synthetic single framework scFv—Application in anti-HDL antibody selections. *New Biotechnol.* **2016**, *33*, 790–796. [[CrossRef](#)]
52. Lowden, M.J.; Henry, K.A. Oxford nanopore sequencing enables rapid discovery of single-domain antibodies from phage display libraries. *Biotechniques* **2018**, *65*, 351–356. [[CrossRef](#)]
53. Barreto, K.; Maruthachalam, B.V.; Hill, W.; Hogan, D.; Sutherland, A.R.; Kusalik, A.; Fonge, H.; DeCoteau, J.F.; Geyer, C. Next-generation sequencing-guided identification and reconstruction of antibody CDR combinations from phage selection outputs. *Nucleic Acids Res.* **2019**, *47*, e50. [[CrossRef](#)]

54. Mei, M.; Zhou, Y.; Peng, W.; Yu, C.; Ma, L.; Zhang, G.; Yi, L. Application of modified yeast surface display technologies for non-Antibody protein engineering. *Microbiol. Res.* **2017**, *196*, 118–128. [[CrossRef](#)]
55. Sha, S.; Agarabi, C.; Brorson, K.; Lee, N.-Y.; Yoon, S. N-Glycosylation Design and Control of Therapeutic Monoclonal Antibodies. *Trends Biotechnol.* **2016**, *34*, 835–846. [[CrossRef](#)]
56. Sydow, J.F.; Lipsmeier, F.; Larraillet, V.; Hilger, M.; Mautz, B.; Mølhøj, M.; Kuentzer, J.; Klostermann, S.; Schoch, J.; Voelger, H.R.; et al. Structure-Based Prediction of Asparagine and Aspartate Degradation Sites in Antibody Variable Regions. *PLOS ONE* **2014**, *9*, e100736. [[CrossRef](#)]
57. Tomar, D.S.; Li, L.; Broulidakis, M.P.; Luksha, N.G.; Burns, C.T.; Singh, S.K.; Kumar, S. In-silico prediction of concentration-dependent viscosity curves for monoclonal antibody solutions. *mAbs* **2017**, *9*, 476–489. [[CrossRef](#)]
58. Obrezanova, O.; Arnell, A.; De La Cuesta, R.G.; E Berthelot, M.; Gallagher, T.R.; Zurdo, J.; Stallwood, Y. Aggregation risk prediction for antibodies and its application to biotherapeutic development. *mAbs* **2015**, *7*, 352–363. [[CrossRef](#)]
59. Sankar, K.; Hoi, K.H.; Yin, Y.; Ramachandran, P.; Andersen, N.; Hilderbrand, A.; McDonald, P.; Spiess, C.; Zhang, Q. Prediction of methionine oxidation risk in monoclonal antibodies using a machine learning method. *mAbs* **2018**, *10*, 1281–1290. [[CrossRef](#)]
60. Mason, D.M.; Friedensohn, S.; Weber, C.; Jordi, C.; Wagner, B.; Meng, S.; Gainza, P.; E Correia, B.; Reddy, S. Deep learning enables therapeutic antibody optimization in mammalian cells by deciphering high-dimensional protein sequence space. *bioRxiv* **2019**.
61. Liu, G.; Zeng, H.; Mueller, J.; Carter, B.; Wang, Z.; Schilz, J.; Horny, G.; Birnbaum, M.E.; Ewert, S.; Gifford, D.K. Antibody Complementarity Determining Region Design Using High-Capacity Machine Learning. *Bioinformatics* **2019**. [[CrossRef](#)]
62. Bujotzek, A.; Lipsmeier, F.; Harris, S.F.; Benz, J.; Kuglstatter, A.; Georges, G. VH-VL orientation prediction for antibody humanization candidate selection: A case study. *Platf. Dev. Expr. Purif. Stable Isot. Labeled Monoclon. Antibodies Escherichia Coli* **2015**, *8*, 288–305. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).