Exploiting next-generation sequencing in antibody selections – a simple PCR method to recover binders

Fortunato Ferrara, Andre A. Teixeira, Leslie Naranjo, M. Frank Erasmus, Sara D’Angelo, and Andrew R. M. Bradbury

ABSTRACT
Antibody discovery using in vitro display technologies such as phage and/or yeast display has become a cornerstone in many research and development projects, including the creation of new drugs for clinical use. Traditionally, after the selection phase, random clones are isolated for binding validation and Sanger sequencing. More recently, next-generation sequencing (NGS) technology has allowed deeper insight into the antibody population after a selection campaign, enabling the identification of many more specific binders. However, this approach only provides the DNA sequences of potential binders, the properties of which need to be fully elucidated by obtaining corresponding clones and expressing them for further validation. Here we present a rapid novel method to harvest potential clones identified by NGS that uses a simple PCR and yeast display approach. The protocol was tested in selections against three different targets and was able to recover clones at an abundance level that would be impractical to identify using traditional methods.

Introduction
Monoclonal antibodies are now widely used for the treatment of many different diseases, including cancer, autoimmunity, infection, migraine and, atherosclerosis. By the end of 2018, there were 87 monoclonal antibodies approved in the US and/or Europe, grossing annual revenues of ~$100B (see ref.1 and www.antibodysociety.org), making monoclonal antibodies the fastest growing drug class, and representing half of all worldwide biopharmaceutical sales. Over 150 biotech and pharmaceutical companies are currently discovering or developing antibody drugs, and the high growth rate is in part due to the higher and faster approval rates for antibodies compared to other drugs. This reflects their straightforward development pathways, high tolerance and relatively low risks of unexpected safety issues in most human clinical trials, immunotherapy excluded.

In vitro antibody selection platforms such as phage and yeast display assist the development of human recombinant antibodies against virtually any desired antigen or epitope that may have arelevant biological function.3 This, in turn, accelerates the discovery process and helps the continuous expansion of the range of diseases that can be treated with these molecules. Naïve phage antibody libraries generated from the B-cells of nonimmune human donors contain antibodies with the potential to recognize all possible antigen types, including human. The technology has been successfully used in the development of U.S.Food and Drug Administration-approved drugs such as adalimumab (AbbVie/Humira*), necitumumab (Eli Lilly/Portrazza*), ramucirumab (Eli Lilly/Cyramza*). During the lead generation phase of drug discovery, antibodies are commonly displayed on the surface of filamentous phage particles either as antigen-binding fragments (Fab) or as a single-chain variable fragment (scFv) domains,4 which consist of the antibody heavy and light chain variable regions connected by linker peptide that preserves the binding properties of the antibody.

When using in vitro display, one of the concerns is that the total diversity of antibodies found in these libraries may not be fully tapped during selections. Theoretical and experimental assessments indicate that one would expect to recover thousands of different antibodies from a106 diversity library, and yet, unless heroic efforts are made,8 the number of antibodies identified is usually less than 50. This is mainly due to two factors:1) many antibody libraries are not as diverse as indicated by the number of transformants,9,10 and2) the way potential binders are traditionally screened identifies leads after extensive individual clone screening, an inefficient approach vexed by the redundancy of abundant clones, and sparse, or absent, representation of most clones that occur in lower frequencies within the selected population. Moreover, clones with high binding activity, or with other prominent functional properties, can be found within the lower frequency clone population. Their rarity can be explained by lower expression levels in apokaryotic system, or by more general biases resulting from the biopanning process.11 An alternative approach is to use next-generation sequencing (NGS), which has massively increased the capacity to sequence millions of clones in every fast and inexpensive way,12 revolutionizing several aspects of biological research, including in vitro and invivo antibody selections.10,13

We routinely use an approach that exploits NGS in in vitro display selection analyses. By combining phage display, yeast display, and NGS, we have shown that the number of
identified antibodies can exceed 1,000, reflecting the far deeper sampling that NGS provides: millions of reads, rather than hundreds of microtiter plate wells screened. This is particularly true if there is bias in the selected population, which is often the case. However, unless antibodies identified by NGS can be easily produced as clones for further analysis, their specificities and properties cannot be assessed, leaving uncertain the true breadth of available reactivities, and risking the loss of many potentially valuable antibodies. The greater the number of different antibodies available, the greater the number of different epitopes, and consequently biological activities, that can be targeted. The caveat of NGS sequencing is that after identifying the desired sequences, one does not have the isolated clones in hand. scFvs are typically composed of ~250 amino acids, and, for large number of potential leads identified after selection campaign, synthesizing their coding sequences can become very costly. Moreover, this approach requires the full sequence of the scFv domain to be known, and most current NGS platforms do not offer sufficiently long reads to provide more than either the VH or VL domain.

Since generating clones from NGS sequence represents the primary bottleneck in the full exploitation of NGS for invitro antibody selection, we and others have pioneered the development of efficient methods to rescue even low abundance scFv sequences identified after NGS analysis. Here, we describe arapid, novel, straightforward method to generate antibody clones identified by NGS that can reach deep into the abundance rank. The method consists of isolating clones of interest identified directly from a selected library output by PCR amplification, where the downstream/reverse primer is complementary to the HCDR3 and JH region of the desired clones, while the upstream/forward primer is clone-independent and anchored within the plasmid (Figure 1). The amplification product is subsequently easily cloned directly into yeast by homologous recombination using the specifically designed yeast-display plasmids, allowing the binding properties of the identified antibodies, such as affinity, to be rapidly measured by flow cytometry.

Results

PDNL6-JH vectors and ScFv recovery strategy

Three JH regions (corresponding to the ones present in the scFv scaffolds of the phage antibody library used for this study) were added to the backbone of the pDNL6 yeast display vector. The chosen sequences correspond to portions of the JH3, JH4, and JH6 human germlines, which are unique, and in the case of JH4, sufficient to recreate the others (JH1, JH2 and JH5) at the amino acid level (Table 1). Each vector contains a BssHII restriction upstream of the JH sequence to be used for vector linearization. Once linearized, each vector can be used to clone an scFv of interest by yeast homologous recombination (Figure 1), once the sequence, and JH gene, is known.

To recover an scFv of interest identified by NGS within a library or selection output, we carry out a PCR using a common forward primer hybridizing to the vector backbone.
upstream of the scFv gene, and long HCDR3-specific primers that anchor within identified HCDR3-JH sequences. In this way the PCR products will have a region at the 5’-end matching the vector backbone and the JH region at the 3’-end is homologous to that present in the acceptor vector. This allows the direct cloning of the scFv of interest by yeast homologous recombination (Figure 1).

**Antibody selections and deep-sequencing analysis**

We tested this novel strategy for rescuing antibodies identified by deep sequencing on the outputs obtained from independent selections on three different targets: B7-H4, CD40 ligand (CD40-L), and OX40. We used an anaíve recombinant scFv library and a procedure consisting of two rounds of phage display selection, followed by two rounds of yeast display sorting, which was previously shown to significantly increase the number of identified target-specific antibodies. As we previously characterized the diversity of the naïve library used in this study by NGS, we were able to analyze the initial frequencies of the enriched HCDR3s, obtained after the three selection campaigns. All the identified HCDR3, including the most abundant ones, showed no over-representation in the initial library, indicating no bias advantage during the screening. The abundances of the selected clones in the naïve library are shown in Table 4. All the clones with reported abundances of < 3.16E-09 were so rare that they were not identified in the data set obtained after sequencing the naïve library. The reason we did not find the rare HCDR3s is that with current sequencing platforms the diversity of an naïve library greatly exceeds the number of sequencing reads, but the generated data are large enough to estimate and evaluate the theoretical diversity of the library.

**ScFv rescue approach**

The isolation strategy of specific clones from selection outputs is depicted in Figure 1. Our approach consisted of rescuing the clones containing: 1) the most abundant HCDR3 sequence, (28.4% abundance for the anti-B7H4 clone, 26.8% for CD40-L and 89.6% for OX40); 2) clones for which HCDR3s are represented at ~5% of the population (5.33% abundance for the anti-B7H4 clone, 6.9% for CD40-L and 3.0% for OX40); 3) ~0.5% of the population (0.52% abundance for the anti-B7H4 clone, 0.50% for CD40-L, while no attempt was performed on OX40 because of the clonal dominance); 4) ~0.1% of the population (0.14% abundance for the anti-B7H4 clone, 0.09% for CD40-L); 5) ~0.05% of the population (0.056% abundance for the anti-B7H4 clone, 0.044% for CD40-L); and 6) ~0.01% of the population (0.015% abundance for the anti-B7H4 clone, 0.015% for CD40-L). These are all shown in Table 2. The goal of testing this broad clone abundance range was to assess, first, the lowest HCDR3 clone abundance that could be effectively isolated using this approach, and second, the lowest abundance at which antibodies reliably recognizing the selection target could be isolated.

Having identified clones from the NGS, we designed oligonucleotides that matched the HCDR3 DNA sequence of each of the clones of interest plus its corresponding JH.
segment. Each primer was 60 nucleotides in length, with 30 nucleotides matching the HCDR3, and 30 nucleotides matching the adjacent JH region. The PCR reactions were performed using the total plasmids prepared from the yeast display selection outputs as templates. To reduce the introduction of unwanted mutations, a high-fidelity polymerase was used in the PCR reactions.

PCR using the designed primers produced fragments of the appropriate size for all abundance ranges. The amplification products were retransformed into the yeast cells along with the appropriate linearized vector containing the corresponding JH sequence. Each of the yeast populations obtained, comprising yeast clones all sharing the same HCDR3, was induced to display scFv on the yeast surface, and the binding of biotinylated antigen was tested by flow cytometry with fluorescently labeled streptavidin. For all cloned outputs except those at the lowest abundance (< 0.02%), a binding population was obtained. For the B7-H4 output obtained by amplifying the 0.015% HCDR3 clonotype, there was some scFv display on the yeast surface, but no target binding, while for CD40-L, the 0.015% population showed no binding and poor antibody display (Figure 3).

Affinity determination of the rescued clones

For each of the scFv-clonotype populations transformed into yeast containing the rescued HCDR3 of interest, and

Table 2. HCDR3 sequences and relative abundances identified for each selection. In bold the HCDR3 we tried to rescue using the described strategy.

| Antigen | % of total | HCDR3 | % of total | HCDR3 |
|---------|------------|-------|------------|-------|
| AKVWGQFDI | 78.429 | AKVWGQGFDI | 15.073 | AKWGLGAFDI |
| AKVWVQRGFDI | 7.865 | AKVWVQRGFDI | 15.073 | AKWGLGAFDI |
| AKGPTVNRGFDI | 5.338 | AKGPTVNRGFDI | 15.073 | AKWGLGAFDI |
| ARHRNRGFDI | 3.126 | ARHRNRGFDI | 15.073 | AKWGLGAFDI |
| ARHSGYRFDI | 3.126 | ARHSGYRFDI | 15.073 | AKWGLGAFDI |
| ARDSRYRGFDI | 1.923 | ARDSRYRGFDI | 15.073 | AKWGLGAFDI |
| ARDSVRGQFDI | 1.923 | ARDSVRGQFDI | 15.073 | AKWGLGAFDI |
| ARDSVRGQFDI | 1.923 | ARDSVRGQFDI | 15.073 | AKWGLGAFDI |

Table 3. Summary of NGS analysis of each selection campaign.

| Antigen | MiSeq Reads | # HCDR3 | Abundance Top_1 | Abundance Top_10 |
|---------|-------------|---------|----------------|-----------------|
| B7H4    | 47829       | 53      | 28.4%          | 77.8%           |
| CD40    | 48401       | 35      | 26.5%          | 82.9%           |
| OX40    | 52766       | 22      | 89.6%          | 94.9%           |
demonstrating binding activity, we sequenced a number of clones to determine whether the rescued population was homogenous or comprised a “mini-library” in which a common HCDR3 was combined with additional mutations elsewhere. We found that, with one exception, clones shared the same HCDR3, but had different point mutations in either one of the other CDRs or in the framework. The one exception had a mutation at the N-terminus of the HCDR3 in one of the clones obtained from the most abundant binder in the CD40-L campaign. Interestingly, but perhaps not surprisingly, such mutations resulted in different affinities. In the case of the clones selected against B7-H4, the antibody affinities ranged from 75 to 310 nM (Table 5), with the most abundant antibodies correlated to the better affinity values, while for CD40-L the range was 92 to 125 nM with no apparent correlation with antibody abundance (Table 5). For the OX40 selection, significant binding activity was observed only for the clones sharing the most abundant HCDR3, with a measured affinity value of 125 nM. Two other rescued HCDR3s had affinities > 1 µM (Table 5). None of the tested antibodies that bound their target showed activity toward an irrelevant negative control, indicating the specificity of all selected antibodies.

For B7-H4, we were able to correctly rescue clones with an HCDR3 abundance as low as 0.1% abundance in the NGS analysis. When we tried to rescue clones at 0.05%, we inadvertently amplified a more abundant clonotype (0.792%) that shared four amino acids at the 3' of the HCDR3 of interest (Figure 4a).

For the CD40-L campaign, the limit of valid clone isolation was 0.497%, since when we tried to rescue clones at an abundance of 0.088% and 0.044% we again amplified more abundant clonotypes (0.792% and 0.044%, respectively) sharing seven and four amino acids at the 3' of the HCDR3 of interest, respectively (Figure 4b).

For the OX40 selection, the limit of valid clone isolation was 0.497%, since when we tried to rescue clones at an abundance of 0.088% and 0.044% we again amplified more abundant clonotypes (0.792% and 0.044%, respectively) sharing seven and four amino acids at the 3' of the HCDR3 of interest, respectively (Figure 4b).

Selection of OX40 provided a very dominant clone with affinity in the nM range. Clonotypes with abundances of 3.04% and 0.59% showed a weak binding signal with affinity above 1 µM (Figures 3 and 4c). For the OX40 selection we decided to also analyze the population of yeast obtained after the first yeast sort, to see if less clonal dominance was present and if it was possible to recover more clones (see Table 6). The results show the most abundant clone in the final yeast sorting was also the most dominant in the previous selection step (from 72.87% to 89.6%), and also these second-highest ranked clone was the same after both sorts with an increase in abundance at the end of the enrichment process (from 1.72% to 3.05%). Two other clones were also shared between the two data sets: one clone showing a slight enrichment (from 0.24% to 0.59%) representing the third-high ranked clone after two sorts and showing minimal binding activity (Figure 3). The fourth one dropped in abundance after thesecond sort, when rescued using the described approach, and no binding activity was detected (data not shown). The remaining clones were present in either of the two data sets (but not both) at very low abundance, probably indicating their presence as “background noise”.

**Using longer primers for rescuing low abundance clones**

The use of 60 base primers was unable to rescue clones with abundances < 0.1%. In order to determine whether primers annealing to the entire HCDR3 sequences of rare clones (plus the clone-specific JH region) could overcome this issue, we carried out an additional set of experiments in which primers were much longer. We focused this analysis on clones we failed to rescue using the 60mer oligonucleotides: the clone with 0.056% abundance obtained during the selection on B7-H4, and the clones obtained from the CD40-L campaign with an abundance of 0.088% and 0.044%. The designed primers ranged from 78 to 87 base pairs. Such long oligonucleotides are more expensive, partly because of their length, but also due to the extra purification required to reduce errors accumulating during the synthesis process. PCR reactions were performed using plasmids prepared from the total yeast display selection outputs obtained after thesecond sort, and the correctly-sized fragments were transformed into yeast, together with the corresponding JH-specific linearized vector. The obtained yeast populations were analyzed for binding by flow cytometry. The population obtained from the B7-H4 clone did not show any binding activity,
reflecting the previously observed correlation between abundance and affinity we saw for this selection campaign, while both populations derived from the CD40-L clonotypes showed specific binding (Figure 5a). For each of the populations transformed into yeast, including the one not showing binding activity, we sequenced four clones. We found that clones with the specific HCDR3s of interest were rescued. For each clonotype, four identical clones were rescued (Figure 5b). The affinities were measured only for the CD40-L clonotypes (Table 5).

Discussion

Display-based technologies, in particular phage display, have been successfully applied to the discovery of novel therapeutic antibodies. The importance of this platform was recognized by the award of the 2018 Nobel Prize to scientists that developed the technology.21 One of the main limitations of the standard screening method is that, during the biopanning process, some binding clones become dominant (as was observed for the OX40 selection), resulting in identification...
of a limited number of different candidates after screening. For example, in our study only 40 of 384 clones would not have been derived from the dominant clone in the OX40 selection. The advent of NGS has dramatically increased the depth and breadth of the analysis of display library selection outputs, making it possible to obtain information not only on the clones identified using a conventional colony screening method, but also additional, less abundant candidates.14,15,20,22 Among the most commonly used NGS platforms, there has been a steady increase in the number and length of reads that can be obtained, but only PacBio has invested in increasing read length to a level commensurate with obtaining full-length sequences of displayed antibody-like molecules, i.e., scFvs and Fabs. For this reason, isolating clones identified by NGS from a selection campaign, to test their functional properties, continues to be a challenge. With PacBio now able to provide full-length clone sequences, gene synthesis is a valuable option, but remains prohibitive on a large scale, even if the cost of gene synthesis is becoming significantly cheaper. Furthermore, gene synthesis usually takes weeks. We previously described an approach allowing the direct isolation of NGS identified clones of interest from a selected library using an inverse PCR23 and ligation-based strategy that exploits the unique barcode nature of the HCDR3 sequence.

Table 5. Affinities and correlations of the individual clones obtained from each clonotype. Clonotype indicated with * represent the ones rescued using longer primers.

| Clones | Affinity Kd (nM) | R² (correlation) | Clones | Affinity Kd (nM) | R² (correlation) | Clones | Affinity Kd (nM) | R² (correlation) |
|--------|------------------|------------------|--------|------------------|------------------|--------|------------------|------------------|
| Clonotype A | #1 95 | 0.9926 | #2 110 | 0.9837 | #3 75 | 0.9882 | #4 78 | 0.9834 | #5 109 | 0.9790 | Clonotype B | #6 102 | 0.9872 | #7 115 | 0.9882 |
| Clonotype B | #1 105 | 0.9911 | Clonotype C | #2 155 | 0.9817 | #3 154 | 0.9802 |
| Clonotype C | #1 175 | 0.9915 | #2 180 | 0.9716 | #3 195 | 0.9793 | #4 178 | 0.9856 | Clonotype D | #1 310 | 0.9827 | #2 310 | 0.9811 |
| Clonotype D | #1 245 | 0.9821 | #2 200 | 0.9802 | #3 195 | 0.9910 | #4 285 | 0.9806 |
| Clonotype E | #1 125 | 0.9811 | #2 125 | 0.9711 | #3 102 | 0.9906 | #4 101 | 0.9794 | Clonotype E* | #1 >1 µM | Not calculated | #2 >1 µM | Not calculated |
| Clonotype E* | #1 155 | 0.9804 | #2 125 | 0.9721 | #3 102 | 0.9902 | #4 92 | 0.9902 | #1 125 | 0.9802 | #2 125 | 0.9802 |

Figure 3. The binding profile of the antibody populations rescued using the present strategy was analyzed by yeast display and flow cytometry analysis. All clones but the ones with 0.015% abundance show expression and binding activity when cloned into the yeast display system.
to synthesize pairs of outward-facing primers that are used to amplify plasmids containing the HCDR3. However, that approach was not very effective for short HCDR3s or clones with abundances less than 1%.

Here, we explored a faster, cheaper method, requiring the design of a single specific primer for each HCDR3 of interest, to quickly translate an NGS analysis into actual clones. We focused our approach on the HCDR3 sequence, since this is...
Table 6. Comparison of the relative abundances of clones present after the first and second yeast sort during the selection campaign against OX40.

| HCDR3 | After first sort Abundance (%) | After second sort Abundance (%) |
|-------|-------------------------------|-------------------------------|
| ARDVGYSGDNWFDP | 72.87 | ARDVGYSGDNWFDP | 89.60 |
| ADRSSGWYSGDMDV | 1.72 | ADRSSGWYSGDMDV | 3.05 |
| ATDRGCSCTSGYVYFDDY | 1.42 | ATDRGCSCTSGYVYFDDY | 0.95 |
| AFRSAYGLDV | 1.11 | AFRSAYGLDV | 0.76 |
| ADRVDVGFWD | 0.79 | ADRVDVGFWD | 0.18 |
| ADRSSGWYSGDMDV | 0.66 | ADRSSGWYSGDMDV | 0.26 |
| ARDSGWYVDPDY | 0.56 | ARDSGWYVDPDY | 0.24 |
| ARSGSYDPFSD | 0.42 | ARSGSYDPFSD | 0.18 |
| ARDSGLGDMDV | 0.37 | ARDSGLGDMDV | 0.14 |
| ARSNSGWGDSD | 0.36 | ARSNSGWGDSD | 0.12 |
| ARAGSWMADDY | 0.34 | ARAGSWMADDY | 0.11 |
| ARGSDWLMDDY | 0.31 | ARGSDWLMDDY | 0.11 |
| ARDSSSWYTSALDV | 0.24 | ARDSSSWYTSALDV | 0.08 |
| ARLPGAMNV | 0.21 | ARLPGAMNV | 0.08 |
| ARTPSSELYLMLDADFY | 0.19 | ARTPSSELYLMLDADFY | 0.07 |
| ARHSSYDDSFADY | 0.18 | ARHSSYDDSFADY | 0.07 |
| ARMRVGGFDY | 0.15 | ARMRVGGFDY | 0.06 |
| AGREYGYYFMD | 0.10 | AGREYGYYFMD | 0.06 |
| VAEGSYSGYATDP | 0.08 | VAEGSYSGYATDP | 0.06 |
| TTEGHSRGSGGYYYYGMDV | 0.04 | TTEGHSRGSGGYYYYGMDV | 0.04 |
| AKDFLDAYSNNYTTYGMDV | 0.04 | AKDFLDAYSNNYTTYGMDV | 0.03 |
| ARHSVHPFADFY | 0.02 | ARHSVHPFADFY | 0.01 |
| AREGYYFGYMDV | 0.02 | AREGYYFGYMDV | 0.01 |
| ARLPSSPADFY | 0.02 | ARLPSSPADFY | 0.01 |
| ATDPWGFYFADFY | 0.01 | ATDPWGFYFADFY | 0.01 |
| AFGSIAARFLG | 0.01 | AFGSIAARFLG | 0.01 |

Figure 5. Analysis of the rescued clones using primers covering the entire HCDR3 sequences of rare clones. (a) Binding profile of the antibody populations rescued using the longer primer strategy analyzed by yeast display and flow cytometry analysis. All clones show expression but binding activity was present only in the clones rescued from the CD40-L campaign. (b) Four clones from each rescued antibody population, including the non-binding one, were Sanger Sequenced. Represented are the HCDR3 used to design our clone specific primers. The scFvs containing the HCDR3s of interest were specifically amplified, including the one with 0.044% of abundance.
99.85% of antibodies containing the same HCDR3 in naïve library do not. This reflects our analysis of both in vivo\textsuperscript{28,29} and invitro\textsuperscript{22} NGS datasets, which indicates that identical HCDR3s can be generated by different VDJ recombination events and are expected to result in very different antibody binding properties.

We demonstrated the approach on three different selection outputs. The selections were performed using a relatively high (100 nM) target concentration in order to preserve diversity and test our working hypothesis with clones at different abundance levels. We were able to isolate and characterize clones present down to 0.1% of the total population obtained by NGS using 60 base primers. Such a threshold is reasonable, considering that in our selection strategy, which combines phage and yeast display, we perform the final steps of the enrichment by sorting 10,000 antigen-binding yeast events. As we were unable to generate binding clones with abundances less than 0.1%, we believe many of these less abundant clones may represent NGS artifacts, although it is likely that rare binders do exist, but are more difficult to isolate. Initially, we limited our primer lengths to 60 bases, in order to optimize the costs of the strategy (longer primers required higher purification standards due to the greater possibility of errors during synthesis). Using 60 base primers we obtained one clone with amplification at the N-terminus of HCDR3 for the most abundant CD40-L binder subpopulation, and when we tried to target clones with an abundance <0.1% using 60mer primers, we obtained more abundant clones that partially shared the targeted HCDR3 sequence. We hypothesize that the use of the Q5 proof-reading polymerase, which has powerful 3’ to 5’ exonuclease activity, to rescue clones of interest from the enriched populations probably allowed primers to be degraded from the 3’ end until they were able to amplify more abundant HCDR3s, which in the case of the least abundant B7H4 clone, resulted in the removal of 18 bases from the 3’ end. The use of anon-proof-reading polymerase (e.g., Taq polymerase) would probably overcome this problem, but at the cost of introducing additional mutations. As an alternate solution, we tested whether longer primers, more stringent for particular HCDR3, would help in the isolation of lower abundance clones. This strategy was successful in the rescue of an clone with 0.044% abundance, that was still able to bind its target (CD40-L) with an affinity similar to the most abundant antibodies in the selection.

This method allows the relatively rapid isolation of clones corresponding to HCDR3s identified by the most common NGS platforms. This requires asingle PCR reaction, no purification, and direct transformation into ayeast display vector thanks to the homologous recombination system. The use of yeast display, in combination with fluorescence-activated cell sorting, also allows further refinement of any particular HCDR3 clone set, by, for instance, generating anpanel of antibodies with similar structure but different affinities. If yeast display and flow cytometry are not available, we believe asimilar approach can also be applied to phage antibody libraries. In this case, after identified antibody genes are amplified, they would be cloned into a corresponding phage display vector using aGibson assembly\textsuperscript{30} (or similar) approach.

The method described here is amenable only for antibodies displayed as scFvs in the VL-VH orientation, with the HCDR3 present at the 3’ end of the molecule. In situation where the molecules are displayed with aVH-VL orientation, the strategy can be optimized around the LCDR3, but, unlike the HCDR3, this is not expected to provide sufficient diversity to distinguish different clones.

Beyond its use for the isolation of NGS identified clones selected from naïve libraries, we believe this approach will also be suitable for other library selections in which NGS is used, such as:1) in error-prone PCR affinity maturation campaigns where preservation of the antibody HCDR3 is important;2) selection from semi-synthetic libraries in which diversity is found in the CDRs embedded within constant scaffolds; or3) in synthetic combinatorial libraries where the CDRs are also synthetically generated. In all these examples the same primer design strategy can be applied to rescue antibodies with HCDR3s of interest, with the advantage that all clones in these libraries will have the same JH region (i.e., of the chosen scaffold for asemi-synthetic library). The power of NGS lies in the ability to analyze the entire depth of an antibody output after aselection campaign. However, the challenge of going rapidly from “sequence to clone” has remained aconsistent bottleneck. Our latest approach to overcome this problem, described here, has clear advantages compared to previous methods. For instance, compared to the previous methodology described by our group and others, based on inverse PCR,\textsuperscript{15,16,31} the present strategy requires the design and synthesis of asingle primer, no purification of the inverse PCR products and subsequent enzymatic ligation to reconstitute aworking plasmid, and no bacterial transformation and subsequent cloning into yeast cells for validation of the binding activity. The isolation limit when using 60 base primers is \textasciitilde0.14%, decreasing to 0.044% when longer primers are used. In both cases binding antibodies could be isolated at these levels, and the lower limit remains to be discovered. The transformation and cloning of PCR products directly into yeast cells by homologous recombination, allows the rapid validation of the system by flow cytometry with further enrichment of higher affinity clones in populations with the same HCDR3. At a practical level, we recommend that 60 base primers are suitable for the isolation of clones >0.2%, but, below this, longer primers that span the whole HCDR3 should be used. When using the 60 base primers, it is important that proposed primer sequences are screened against the whole NGS output to ensure that more abundant similar clones are not amplified, bearing in mind that up to 18 bases may be removed by the polymerase exonuclease activity.

The direct transformation of yeast cells with adisplay vector and PCR products obtained with auniversal primer specifich to the heavy chain CDR3 and framework 4 regions of the scFv sequence is here shown to be an effective methodology for the recovery of scFvs, even for clones at very low abundance clones within aselection output.
Materials and methods

Bacterial and yeast strains

Omnimax™ (ThermoFisher Scientific, C854003): F′[proAB lacI8 lacZΔM15 7n10(TetR) Δ(ccdAB) mcrAΔ(mrr hsdRRMS-mcrBC) Φ 80(lacZ)ΔM15 Δ(lacZYA-argF) U169 endA1 recA1 supE44 thi-1 gyrA96 relA1 tonApanD

EBY100 (ATCC®, MYA-4941): MATa AGA1+::GAL1- AGA1-::URA3 ura3-52 trp1 leu2delta200 his3- delta200 pep4::HIS3 prb11.6R can1 GAL

Generation of pDNL6-JH vectors

We generated three different vectors to match the terminal part of the JH sequences present in our library: JH3 (GTMVTVSS), JH4 (GTVLTVSS) and JH6 (GTVTVTVSS). The JH1, JH2 and JH5 germlines share the same amino acid sequence with JH4, making them compatible with the vectors. Synthetic oligo nucleotides (Genewiz, NJ) encoding the two strands of each JH regions

(Table 1) were annealed and ligated to the pDNL6 vector previously digested with the enzyme.

We transformed the ligation products into E.coli Omnimax™ (ThermoFisher Scientific, C854003) by heat shock and single colonies were sequenced to confirm the presence of the desired sequence.

ScFv antibody selections

The antigens used for the scFv selections were B7-H4 (ACRObiosystem, B74-H5222), CD40-L (Shenandoah Biotechnology, 100-25AF), and OX40 (ACRObiosystem, OXO-H5224). The targets were biotinylated using the EZ-Link NHS-LC-LC-Biotin system (Thermo Fisher Scientific), following the manufacturer’s instructions. The phage naïve scFv library, was used for two rounds of phage display against the antigens followed by two rounds of yeast display sorting. The detailed protocols for antibody selections against biotinylated proteins are described in Ferrara etal. 14,18

Deep sequencing

Plasmid DNA was extracted from the final yeast populations, specifically enriched for their binding to the antigens of interest, using the QIAprep spin miniprep kit (Qiagen, 27104), adding 50–100 µl of acid-washed glass beads (Sigma, G-8772) during the resuspension step and vortexing for 5 min before proceeding with the miniprep protocol.

The plasmid DNA was used directly as template for NGS sample preparation, which consisted of PCR amplification of the VH domains of the selected scFv. The designed primers contain different barcodes that allowed us to pool different amplicons together in order to ligate Illumina adapters to asingle pool. The PCR amplification was performed using Q5 high fidelity polymerase (NEB, M0491) following the manufacturer’s protocol. PCR products were gel purified, quantified and sent to Genewiz™ for sequencing using their Amplicon-EZ service. This is a simpler, accelerated version of MiSeq NGS, providing a fast and convenient way to sequence up to 50,000 reads per sample.

Briefly, for the bioinformatic analysis, all the paired-end sequence reads derived from MiSeq runs were assembled using PEAR 32 and the Vregion germlines and CDRs were annotated using the IgBlast command-line tool. After appropriate alignment, all HCDR3 amino acid sequences were aligned in pairs using the Needleman–Wunsch global alignment algorithm and the PAM30 substitution matrix,33 which assigned score to each pair, the higher the score the higher the similarity between the HCDR3 sequences. Subsequently, the alignment scores were stored in asquare matrix and clustered using the DBSCAN algorithm (e = 135),34 which created groups (clusters) based on the similarity between the different HCDR3 sequences.

Primer design and PCR amplification

From the plasmid preps obtained from the final yeast-sorted populations, specific scFv clones containing the HCDR3 of interest were amplified using ahigh-fidelity polymerase (Q5 Polymerase, NEB, M0491) and 0.1 ng of template DNA, which represents 100–1000 times the diversity of the selection output. The 5’ primer consist of a“universal” primer that anneals to the yeast display vector with enough homology for the subsequent homologous recombination of the PCR product into the pDNL6-JH vector of choice, determined by the JH sequence of the scFv of interest. The 3’ primers were designed on the DNA consensus sequence for the HCDR3s of interest: each HCDR3 represent acluster consisting of merged, highly homologous, HCDR3s; the most frequent sequence was used as the consensus sequence within acluster to design the primers employed in our strategy. The first set of 3’ primer length was limited to 60bp, with the primer matching the identified JH region of the scFv of interest and the rest of the oligonucleotide consisting of the consensus sequence of the targeted HCDR3. Longer 3’ primers were designed matching the identified JH region of the scFv of interest and entire consensus sequence of the HCDR3 of interest.

Yeast transformation and sequencing

The partially amplified scFvs were sub-cloned into the pDNL6-JH yeast display vector as previously described, this was possible thanks to the overlapping regions introduced by PCR: within the vector upstream of the scFv cloning site, and in the JH region of the antibody, which is also present in each specific pDNL6-JH vector. The vector and the fragments were cotransformed into yeast cells to allow cloning by gap repair.35

To test the binding of the recovered scFvs, the transformed yeast cells were induced and 1 × 10⁶ cells were stained with 100 nM of biotinylated antigen. Subsequently, the cells were labeled with anti-SV5-PE to detect the scFv display level on the cell surface and streptavidin-AlexaFluor633 to detect the antigen bound by the expressing scFvs.
**Amplicon sequencing and sanger sequencing**

After testing the binding activity of the transformed cells, plasmid DNA was extracted from the yeast cells and used as template to prepare PCR products to be analyzed by NGS, where we obtained approximately 50,000 sequences per sample, enough to prove the efficacy of our scFv rescuing strategy. Data was analyzed as described before. In parallel, the same plasmid preps were used to transform competent bacterial cells to obtain single clone sequences and to obtain single clones plasmid DNA to be transformed back into yeast if necessary.

**Binding assay**

Taking advantage of the flexibility of flow cytometry, we performed antigen dose response experiments on single yeast clones to determine the affinity of the displayed scFv, without subcloning, expression and purification of the recombinant antibody, following published methods.36

**Disclosure of potential conflicts of interest**

No potential conflicts of interest were disclosed.

**Funding**

This work was supported by the National Institutes of Health under Grant R43CA224843 awarded to ARMB; National Institutes of Health [1R43CA224843-01].

**ORCID**

André A. Teixeira http://orcid.org/0000-0002-8348-0235
Leslie Naranjo http://orcid.org/0000-0003-2165-0814
M. Frank Erasmus http://orcid.org/0000-0002-8046-5048
Sara D’Angelo http://orcid.org/0000-0002-8329-9938

**References**

1. GriloAL, MantalarisA. The increasingly human and profitable monoclonal antibody market. Trends Biotechnol. 2019;37:9–16. doi:10.1016/j.tibtech.2018.05.014.

2. KaplounH, ReichertJM. Antibodies to watch in 2019. Mabs. 2019;11:219–38. doi:10.1080/19420862.2018.1556465.

3. BradburyAR, Sidhus, Dubels, McCaffertyJ. Beyond natural antibodies: the power of in vitro display technologies. Nat Biotechnol. 2011;29:245–54. doi:10.1038/nbt.1791.

4. HustonJS, LevinsonD, Madgett-HunterM, TaiMS, NovotnyJ, MargoliesMN, RidgeRJ, BrucoleriRE, HaberE, CrealR, et al. Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. Proc Natl Acad Sci USA. 1988;85:5879–83. doi:10.1073/pnas.85.16.5879.

5. PerelsonAS, OsterGF. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-non-self discrimination. JTheor Biol. 1979;81:645–70. doi:10.1016/0022-5193(79)90275-3.

6. GriffithsAD, WilliamsSC, HartleyO, TomlinsonMW, WaterhouseP, CrosbyWL, KontermannRE, JonesPT, LowNM, AllisonJ, et al. Isolation of high affinity human antibodies directly from large synthetic repertoires. Embo J. 1994;13:3245–50. doi:10.1002/embi.199413issue-14.

7. MarksJD, HoogenboomHR, BonnerTP, McCaffertyJ, GriffithsAD, WinterG. Bypassing immunization. Human antibodies from V-gene libraries displayed on phage. J Mol Biol. 1991;222:581–97. doi:10.1016/0022-2836(91)90498-U.

8. EdwardsBM, BarashSC, MainSH, ChoiGH, MinterR, UllrichS, WilliamsE, Du Foul, WiltonJ, AlbertVR, et al. The remarkable flexibility of the human antibody repertoire; isolation of over one thousand different antibodies to asingle protein, BLyS. J Mol Biol. 2003;334:103–18. doi:10.1016/j.jmb.2003.09.054.

9. Glimmikelj, D’AngeloS, KhanTA, ReddyST, NaranjoL, FerraraF, et al. Deep sequencing in library selection projects: what insight does it bring? Curr Opin Struct Biol. 2015;33:146–60. doi:10.1016/j.sbi.2015.09.001.

10. Glimmikelj, ZhaiW, Berkal, TelmanD, HuertaG, MehtaGR, Nili, Meil, SundarPD, DayGMR, et al. Precise determination of the diversity of acombinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc Natl Acad Sci USA. 2009;106:20216–21. doi:10.1073/pnas.0909775106.

11. YangW, YoonA, LeeS, KimS, HanJ, ChungJ. Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. Exp Mol Med. 2017;49:e308. doi:10.1038/emm.2017.22.

12. NiedringhausTP, MilanovaD, KerbyMB, SnyderMP, BarronAE. Landscape of next-generation sequencing technologies. Anal Chem. 2011;83:4327–41. doi:10.1021/ac2010857.

13. GeorgiouG, IppolitoGC, Beausangl, BasseCE, WardemannH, QuakeSR. The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat Biotechnol. 2014;32:158–68. doi:10.1038/nbt.2782.

14. FerraraF, D’AngeloS, Gaiottoti, NaranjoL, TianH, GraslundS, DobrevetskyE, HaberP, Lund-JohansenF, SaragozzaS, et al. Recombinant renewable polyclonal antibodies. Mabs. 2015;7:32–41. doi:10.1089/mabs.2014.0862.2015.598047.

15. D’AngeloS, KumarS, NaranjoL, FerraraF, KissC, BradburyAR. From deep sequencing to actual clones. Protein Eng Des Sel. 2014:27:301–07. doi:10.1093/protein/gzn032.

16. SpiliotopoulosA, OwenJP, MaddisonBC, Drenvelny, ReesHC, GoughKC. Sensitive recovery of recombinant antibody clones after their in silico identification within NGS datasets. JImmunol Methods. 2015;420:50–55. doi:10.1016/j.jim.2015.03.005.

17. SblatteroD, BradburyA. Exploiting recombination in single bacteria to make large phage antibody libraries. Nat Biotechnol. 2000;18:75–80. doi:10.1038/71958.

18. FerraraF, NaranjoLA, KumarS, Gaiottoti, MukundanH, SwansonB, et al. Using phage and yeast display to select hundreds of monoclonal antibodies: application to antigen 85, atuberculosis biomarker. PLoS One. 2012;7:e49535. doi:10.1371/journal.pone.0049535.

19. D’AngeloS, StaquicinIF, FerraraF, StaquiciniDI, SharmaG, TarletonCA, NguyenH, NaranjoLA, SidmanRL, ArapW, et al. Selection of phage-displayed accessible recombinant targeted antibody (SPARTA): methodology and applications. JCI Insight. 2018;3.doi:10.1172/jci.insight.98305.

20. FerraraF, NaranjoLA, KumarS, Gaiottoti, MukundanH, SwansonB, et al. The antibody mining toolbox: an open source tool for the rapid analysis of antibody repertoires. Mabs. 2014;6:160–72. doi:10.4161/mabs.27105.

21. ChanCE, LimAP, MacAryPA, HansonBJ. The role of phage display in therapeutic antibody discovery. IntImmunol. 2014;26:649–57. doi:10.1055/s-0034-1381082.

22. D’AngeloS, FerraraF, NaranjoL, ErasmusMF, HaberP, BradburyARM. Many routes to an antibody heavy-chain CDR3: necessary, yet insufficient, for specific binding. Front Immunol. 2018;9:395. doi:10.3389/fimmu.2018.00395.

23. HoskinsRA, StapletonM, GeorgeRA, YuC, WanKH, CarlsonJW. Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP). Nucleic Acids Res. 2005;33:e185. doi:10.1093/nar/gni184.

24. BarbacisCF 3rd, BajajD, HoekstraDM, LernerRA. Semisynthetic combinatorial antibody libraries: achemical solution to the diversity problem. Proc Natl Acad Sci USA. 1992;89:4457–61. doi:10.1073/pnas.92.9.4457.

25. BraunagelM, LittleM. Construction of asemisynthetic antibody library using trinucleotide oligos. Nucleic Acids Res. 1997;25:4690–91. doi:10.1093/nar/25.22.4690.
26. MahonCM, LambertMA, GlanvilleJ, WadeJM, FennellBJ, KrebsMR, ArmellinoD, YangS, LiuX, O’SullivanCM, et al. Comprehensive interrogation of a minimalist synthetic CDR-H3 library and its ability to generate antibodies with therapeutic potential. J Mol Biol. 2013;425:1712-30. doi:10.1016/j.jmb.2013.02.015.

27. XuJL, DavisMM. Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. Immunity. 2000;13:37–45. doi:10.1016/S1074-7613(00)00006-6.

28. DeKoskyBJ, LunguOI, ParkD, JohnsonEL, CharabW, ChrysostomouC, KurodaD, EllingtonAD, IppolitoGC, GrayJJ, et al. Large scale sequence and structural comparisons of human naive and antigen-experienced antibody repertoires. Proc Natl Acad Sci USA. 2016;113:E2636–45. doi:10.1073/pnas.1525510113.

29. DeWittWS, LindauP, SnyderTM, SherwoodAM, VignaliM, CarlsonCS, GreenbergPD, DuerkoppN, EmersonRO, RobinsHS, et al. A public database of memory and naive B-cell receptor sequences. PLoS One. 2016;11:e0160853. doi:10.1371/journal.pone.0160853.

30. GibsonDG, YoungL, ChuangRY, VenterJC, HutchisonCA 3rd, SmithHO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6:343–45. doi:10.1038/nmeth.1318.

31. FerraraF, BradburyARM, D’AngeloS. Primer design and inverse PCR on yeast display antibody selection outputs. Methods Mol Biol. 2018;1721:35–45.

32. ZhangJ, KoberbK, FlouriT, StamatakisA. PEAR: a fast and accurate Illumina Paired End reads. Bioinformatics. 2014;30:614–20. doi:10.1093/bioinformatics/btt593.

33. NeedlemanSB, WunschCD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol. 1970;48:443–53. doi:10.1016/0022-2836(70)90057-4.

34. EsterM, KriegelH-P, SanderJ, XuX. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD’96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; Portland, OR; 1996:226–31.

35. BoderET, WittrupKD. Optimal screening of surface-displayed polypeptide libraries. Biotechnol Prog. 1998;14:55–62. doi:10.1021/bp970144q.

36. FeldhausMJ, SiegelRW, OpreskoLK, ColemanJR, FeldhausJM, YeungYA, CochranJR, HeinzelmanP, ColbyD, SwersJ, et al. Flow-cytometric isolation of human antibodies from an immunoblot Saccharomyces cerevisiae surface display library. Nat Biotechnol. 2003;21:163–70. doi:10.1038/nbt785.