Journal of Immunological Methods xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Technical note

# Barcoded sequencing workflow for high throughput digitization of hybridoma antibody variable domain sequences

Yongmei Chen[a], Si Hyun Kim[a], Yonglei Shang[a], Joseph Guillory[c], Jeremy Stinson[c], Qing Zhang[a,b], Isidro Hötzel[a], Kam Hon Hoi[a,b,*]

[a] Department of Antibody Engineering, Genentech Inc., South San Francisco, CA 94080, United States
[b] Department of Bioinformatics and Computational Biology, Genentech Inc., South San Francisco, CA 94080, United States
[c] Department of Molecular Biology, Genentech Inc., South San Francisco, CA 94080, United States

ARTICLE INFO

*Keywords:*
Antibody sequencing
Next Generation Sequencing
Hybridoma sequencing
Bioinformatics workflow
Barcoded primers
96-Well plate

ABSTRACT

Since the invention of Hybridoma technology by Milstein and Köhler in 1975, its application has greatly advanced the antibody discovery process. The technology enables both functional screening and long-term archival of the immortalized monoclonal antibody producing B cells. Despite the dependable cryopreservation technology for hybridoma cells, practicality of long-term storage has been outpaced by recent progress in robotics and automations, which enables routine identification of thousands of antigen specific hybridoma clones. Such throughput increase imposes two nascent challenges in the antibody discovery process, namely limited cryopreservation storage space and limited throughput in conventional antibody sequencing. We herein provide a barcoded sequencing workflow that utilizes next generation sequencing to expand the conventional sequencing capacity. Accompanied with the bioinformatics tools we describe, the barcoded sequencing workflow robustly reports unambiguous antibody sequences as confirmed with Sanger sequencing controls. In complement with the commonly accessible recombinant DNA technology, the barcoded sequencing workflow allows for high throughput digitization of the antibody sequences and provides an effective solution to the limitations imposed by physical storage and sequencing capacity.

## 1. Introduction

The biotechnology industry has observed a tremendous growth at double-digit rate in recent years (Aggarwal, 2014); among the biologics, close to 40% of the US sales were contributed by the monoclonal antibodies therapeutics. The success of the monoclonal antibodies sector would not have been possible without the hybridoma technology invented by Milstein and Köhler in 1975 (Köhler and Milstein, 1975). Ribatti's review (Ribatti, 2014) of the hybridoma technology provided a detailed history and its significance in therapeutic application. Briefly, the technology entails the fusion of mouse myeloma cells with primary B cells derived from immunized animals, generating immortalized hybrid cells known as hybridomas. The fusion with myeloma cells allows the primary B cell to sustain monoclonal antibody secretion beyond its usual lifespan thus enabling monoclonal antibody secretion to sufficient amounts for functional screening. Additionally, unhindered proliferation of hybridomas enabled sub-sampling of the expanded hybridomas for archival purpose, as well as for immunoglobulin gene sequencing without interrupting antibody production and screening purposes. Since the inception of the technology, many commercially successful monoclonal antibody biologics have been discovered through hybridoma technology (Ribatti, 2014). In fact, hybridoma technology in the antibody discovery process is still extremely effective and extensively used in the biotechnology industry albeit some low-throughput bottlenecks that necessitate a high throughput retrofit.

One such bottleneck involves the identification of the monoclonal antibody sequences. Typically, there is a minimum requirement of 20 ng of DNA for each variable chain of the hybridoma's antibody that is required by the Sanger sequencing method (Sanger et al., 1977). Therefore, individual polymerase chain reaction needs to take place for each chain in the hybridoma clone. While the use of capillary electrophoresis and automation techniques in the past decades have greatly improved the Sanger sequencing throughput (Smith et al., 1986), the state-of-the-art instrument remains limited at a capacity of sequencing 96 or 384 PCR samples simultaneously with a 24 h turnaround. From the perspective of sequencing throughput, it is rather impractical to scale-up the number of Sanger sequencers to parallel the throughput in the order of tens of thousands. And, from the perspective of sequencing

* Corresponding author at: Department of Antibody Engineering, Genentech Inc., South San Francisco, CA 94080, United States.
*E-mail address:* hoi.kam-hon@gene.com (K.H. Hoi).

https://doi.org/10.1016/j.jim.2018.01.004
Received 8 August 2017; Received in revised form 10 January 2018; Accepted 16 January 2018

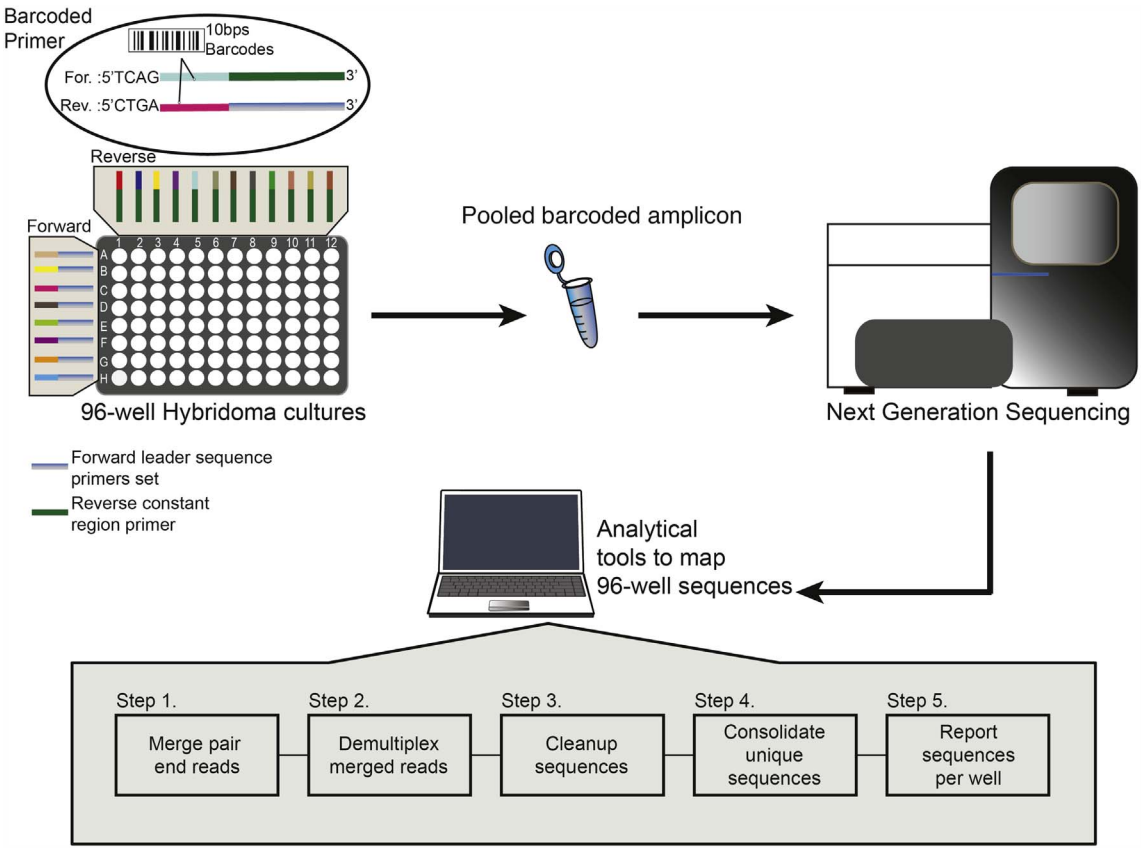


**Fig. 1.** Schematic overview of the barcoded sequencing workflow. The cDNAs for each of the Hybridoma clones in each of the 96-well are generated and purified. Pooled amplicons are sequenced using Illumina MiSeq sequencer with the 2 × 300 bps chemistry platform. Bioinformatics tools are used to identify and map the pooled sequences to its corresponding wells.

efficiency, Sanger sequencing is intolerable to impurities in the input DNA materials. In other words, if the input DNA materials were derived from, for example, a polyclonal pool of hybridomas, or a hybridoma expressing multiple B-cell-derived heavy or light chains, or a hybridoma expressing a myeloma endogenous light chain, the sequencing read quality and yield would be unfavorably affected leading to uninformative sequencing results. In order to resolve to informative sequencing results, it would be necessary to sub-clone the source DNA materials to deconvolute to monoclonal DNA level prior to re-sequencing. This laborious process would extend the turnaround time by a minimum of 2 to 3 days before informative sequencing results could be obtained. Taken collectively, the lack of efficient and high-throughput antibody sequencing methods is stalling further throughput increase in the hybridoma antibody discovery process. Given the throughput-limiting step of conventional sequencing, it would become apparent that cryopreservation storage space for staging the hybridoma backlogs would render another logistical bottleneck for the antibody discovery process.

With the advent of next generation sequencing (NGS) technology, there is an increasing shift from automated Sanger sequencing to the adoption of NGS in the field of biology and medicine (Soon et al., 2013). A detailed background review of the NGS technology can be found in Metzker's article (Metzker, 2010). Briefly, the NGS technology is a culmination of breakthroughs in microfluidics, sequencing chemistry, and imaging/detection techniques that enables massively parallelized sequencing of up to billions of DNA molecules. Through the mechanism of sequencing-by-synthesis, most NGS sequencers achieved high throughput sequencing via monitoring the incorporation of the nucleotides during DNA synthesis from the numerous template amplicons simultaneously. With the power of PCR amplification, any trace amount of DNA could be amplified to form sequencing amplicons resulting in NGS sequencing reads. This high sensitivity enables the detection of minuscule amount of DNA where it would normally be undetectable using Sanger sequencing. Specifically, our experience using spike-in controls with NGS (Illumina MiSeq) had been consistently yielding on average 12 reads with deviations of 4 reads for a 10 pg spike-in control, of which Sanger sequencing was not able to detect. In addition to sensitivity, each Illumina MiSeq sequencing run could generate as many as 25 million reads of heavy or light chain variable domain sequences. Due to the high sequencing depth, financially speaking, it is more economical to use NGS when sequencing large number of samples. In terms of costs for sequencing, Fig. S1 showed a direct comparison between Sanger sequencing and NGS indicating a cost benefit of NGS when sequencing large amount of samples, especially when sequencing > 6 plates of 96-well plate samples.

The unprecedented sequencing throughput and sensitivity allow for broad applications (Shendure and Aiden, 2012). Specifically in the field of antibody, it has been utilized to better understand antibody repertoire (Benichou et al., 2012; Calis and Rosenberg, 2014; Georgiou et al., 2014; Robinson, 2014); as well as, it has been used for antibody engineering and discovery (Reddy et al., 2010; Koenig et al., 2015). Notably, Wardemann's group described an NGS approach on using single-cell PCR together with 2 dimensional barcoded primer matrix to sequence index-sorted single cells (Busse et al., 2014). This approach was shown to be efficient at producing sequencing results at the single cell level in wells. We found this approach to be quite applicable to the hybridoma antibody discovery workflow, especially with the use of barcoded primers to facilitate pooled sequencing of cells from multiple wells. However, the hybridoma antibody discovery workflow presents its unique set of challenges such as endogenous light chain and multiple B-cell-derived heavy or light chains described above; hence, Wardemann's approach could not be applied directly due to its intended design to detect at single cell level with single heavy and single light chain transcripts. To the best of our knowledge, we have not seen literature