Evaluation of B cell receptor profiling platforms and analysis tools using next-generation sequencing data.

Lily Zheng[1], Yong Zhang[1], Ronit Mazor[2], Chen Song[3], Ariel Erijman[3], Bradley Langhorst[3], Pingfang Liu[3], Andrew Barry[3], Douglas Zhang[4], Andrew Farmer[5], Shaveta Goyal[5], Andrea Baines[1], Lorraine Pelosof[1], Mark Izraelson[6], Stan Ploskavsky[6], Dmitriy Bolotin[6], Sam Westreich[7], Ben Busby[7], Stanley Lan[7], Omar Serang[7], Elaine Johanson[7], Julia Beaver[1], Wenming Xiao[1]

[1]CDER, [2]CBER, [3]New England Biolabs, [4]AbHelix, [5]Takara Bio, [6]MiLaboratories, [7]PrecisionFDA

Abstract

The rapid emergence of novel analysis methods surrounding the adaptive immune system and advancements in immunotherapy and drug development has created the need to standardize the quantitative study of human immune cell receptor repertoire composition and diversity. This research will elucidate current limitations, provide reference data sets, and establish best practices for reconstructing B-cell receptor (BCR) repertoires from sequencing data. Phase I of this study consists of characterization of 50 B-cell cell lines and a Pilot Study assessing 3 BCR-seq assays. We characterized cell lines using BCR-seq, RNA-seq, and whole genome sequencing to determine suitable cell lines for mixing into contrived samples for the full study. BCR repertoire analyses of 50 cell lines across four sequencing technologies reveal clonality and major clonotypes, with good consistency across most cell lines. In the Pilot Study, contrived samples created with two cell lines and PBMCs were sequenced using 3 BCR assays: AbHelix, New England Biolabs (NEB), and Takara Bio. Analyses with MiXCR revealed unexpected clonotype percentages and ratios, which may be partially explained by the BCR expression abundances of the cell lines and PBMCs. This highlights the importance of thorough cell line characterization prior to creating contrived samples.

Introduction

High-throughput sequencing of B-cell receptor (BCR) gene regions and downstream analysis tools have paved the way for a more complete and comprehensive description of B-cell repertoire composition and diversity. Although this field has made vast strides over the last ten years, there are clear inadequacies and shortcomings in the current, downstream analysis methods and their utilization for efficacy determination.

Study Objectives

- Elucidate current limitations
- Provide reference samples and data sets
- Establish best practices for reconstructing BCR repertoires from next-generation sequencing data

Significance

This study will…

- Improve our technical understanding of BCR profiling and thus immune cell function and B-cell adaptive immunity.
- Advance our ability to provide orthogonal, functional evidence for the efficacy of cancer therapies such as immunotherapies and vaccines.
- Allow for the comprehensive evaluation of immunological disorders, cancer treatments, pathogenic response, germline-targeted vaccine design, immunotherapies, and drug efficacy.

We will generate a series of contrived reference samples, using well-characterized B-cell lines, to benchmark the performance of BCR-Seq assays, sequencing, and informatics tools. In doing so, we will establish the technical standards and protocols for reliable and accurate BCR profiling.

Materials and Methods

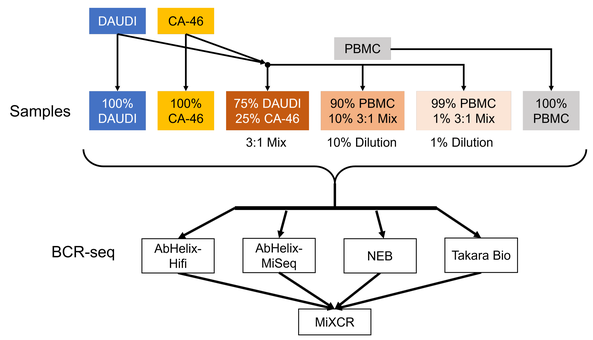

Figure 1. Pilot Study design. Fresh cultured cells from DAUDI and CA46 were mixed at 3-to-1 ratio and diluted into a PBMC pool at 10% and 1% per cell counts. The resulting 6 samples were sequenced using BCR-seq assays from AbHelix, New England Biolabs (NEB), and Takara Bio and analyzed using MiXCR.

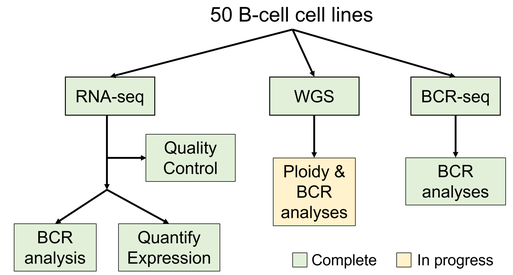

Figure 2. Characterization of 50 B-cell cell lines. We characterized 50 B-cell cell lines using 3 sequencing methods. First, RNA sequencing was performed to determine BCRs and their expression abundance. Second, whole-genome sequencing (WGS) was performed to confirm BCR sequences and determine ploidy of the cell lines. Third, BCR-seq was performed to confirm BCR clonotypes of the cell lines.

Results

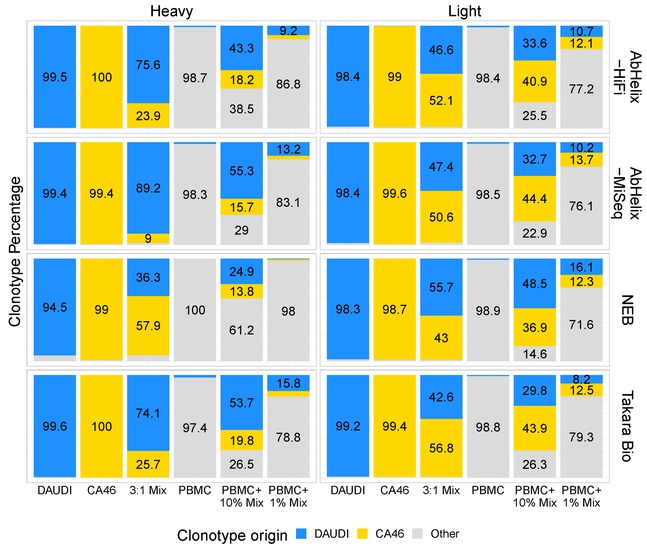

Figure 3. Clonotype distributions for heavy and light chains across Pilot Study samples. Clonotypes were identified by MiXCR v4.2.0-26-develop. Major clonotypes found in DAUDI and CA46 are highlighted in blue and yellow, respectively.

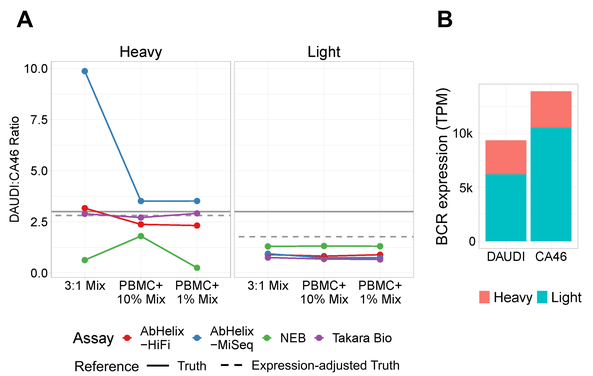

Figure 4. DAUDI to CA46 clonotype ratios and expression abundance of heavy and light chain transcripts. (A) Ratio of DAUDI and CA46 clonotypes identified in samples with 3:1 DAUDI to CA46 mix. **(B)** Abundance of heavy and light chain transcripts represented in transcripts per million (TPM) determined from RNA sequencing.

Results continued

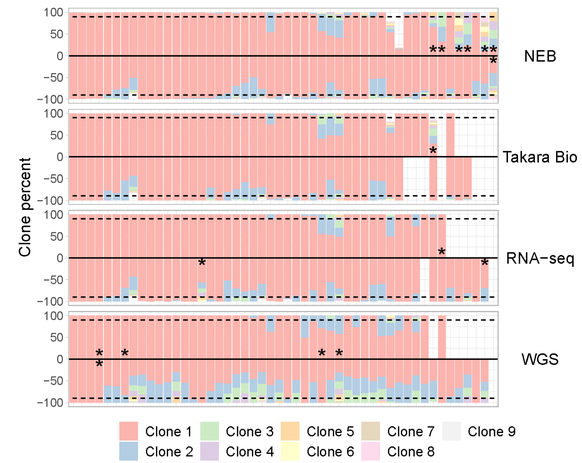

Figure 5. Distributions of top 9 clones across 50 B-cell cell lines. Clonotypes identified by MiXCR are colored to visualize clonality in heavy chain (positive y-axis) and light chain (negative y-axis). Each column contains results for a single cell line. The asterisk (*) indicates low coverage or number of unique UMIs. For some cell lines, unexpressed BCR clonotypes can be observed from WGS.

Discussion and Future Direction

- Overall, results from clonotype analyses across 4 sequencing assays were largely consistent. For some cell lines, unexpressed BCR clonotypes were observed from WGS data. From the 50 cell lines, we selected 9 monoclonal cell lines with different heavy chain V-genes for the first batch of the full study.
- Differences between observed and expected clonotype percentages and ratios of DAUDI:CA46 may be partially explained by the relative BCR expression abundances of the cell lines and PBMCs. Expression abundance of cell lines and PBMCs must be accounted for when creating reference samples for the full study.
- Contrived reference samples for the full study will be created using two different mixing strategies: by cells and by RNA/DNA content.

Acknowledgements

This project was supported in part by an appointment to the Research Fellowship Program at the Office of New Drugs/Center for Drug Evaluation and Research, U.S. Food and Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and FDA.