Abstract #36 | Poster

The Clinico-Genomic Lymphoma Dataset Explorer (CLYDE): A Comparative Analysis of High-Throughput DNA Sequencing Datasets of Diffuse Large B-Cell Lymphoma

J. Zhou[1, 2], F. Ullrich[1], M. Nezamabadi[1, 2], C. Reinhardt[1], B. von Tresckow[1], E. Kocakavuk[1, 2]

[1] *University Hospital Essen, Dept of Hematology and Stem Cell Transplantation,West German Cancer Center, NCT-West, Essen, Germany*
[2] *University Hospital Essen, Institute for AI in Medicine (IKIM), Essen, Germany*

Background: Diffuse large B-cell lymphoma (DLBCL), the most common adult non-Hodgkin lymphoma, exhibits substantial clinical, pathological, and molecular heterogeneity across published cohorts. Despite this diversity, there is currently no comprehensive overview that provides comparisons across molecular studies.

Methods: We conducted a systematic search of PubMed, EMBASE, Web of Science, Cochrane, and ClinicalTrials.gov for human DLBCL studies that reported original high-throughput DNA sequencing data up to October 20, 2025. We excluded studies without original DNA sequencing data, non-human studies, and case reports with fewer than five patients. To facilitate cross-cohort analysis, we developed CLYDE, an interactive cBioPortal application that standardizes core clinical variables and enables side-by-side cohort comparisons, automated statistical testing of clinical distributions, subgroup survival analyses, and visualization of patient selection across datasets.

Results: Of the 1,976 records examined, 288 studies were eligible for inclusion, encompassing a total of 32,594 patients. The number of sequencing studies increased nonlinearly over time, with two pronounced stepwise surges and clear inflection points around 2016 and 2021. Targeted sequencing was the most frequently used approach, employed in the majority of cohorts (212/288 [73.6%]). Most studies were small (157/288 [54.5%] enrolled ≤50 patients), and the number of studies declined sharply with increasing cohort size. Across representative studies, substantial heterogeneity was observed in patient-level characteristics and survival endpoints.

Conclusions: We provide a systematic map of published DLBCL DNA sequencing studies, highlighting clinico-genomic heterogeneity across cohorts. With CLYDE, we have created an interactive tool that allow for easy browsing of datasets and supports cross-cohort validation for future DLBCL sequencing research.

Keywords: Diffuse large B-cell lymphoma (DLBCL), High-throughput DNA sequencing, Clinico-genomic integration, cBioPortal, Cross-cohort heterogeneity

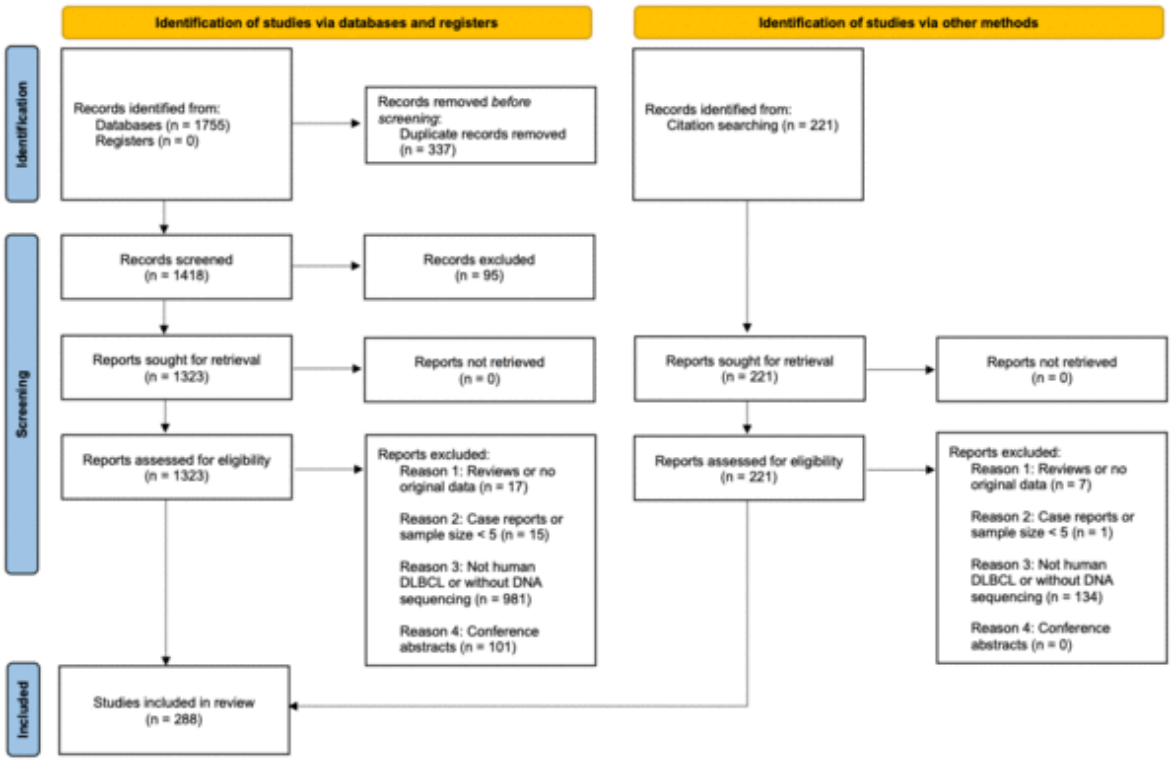

Figure 1. Study Selection Flow Diagram

Flowchart detailing the systematic literature search and screening process, resulting in the inclusion of 288 eligible high-throughput DNA sequencing studies from 1,976 screened records.

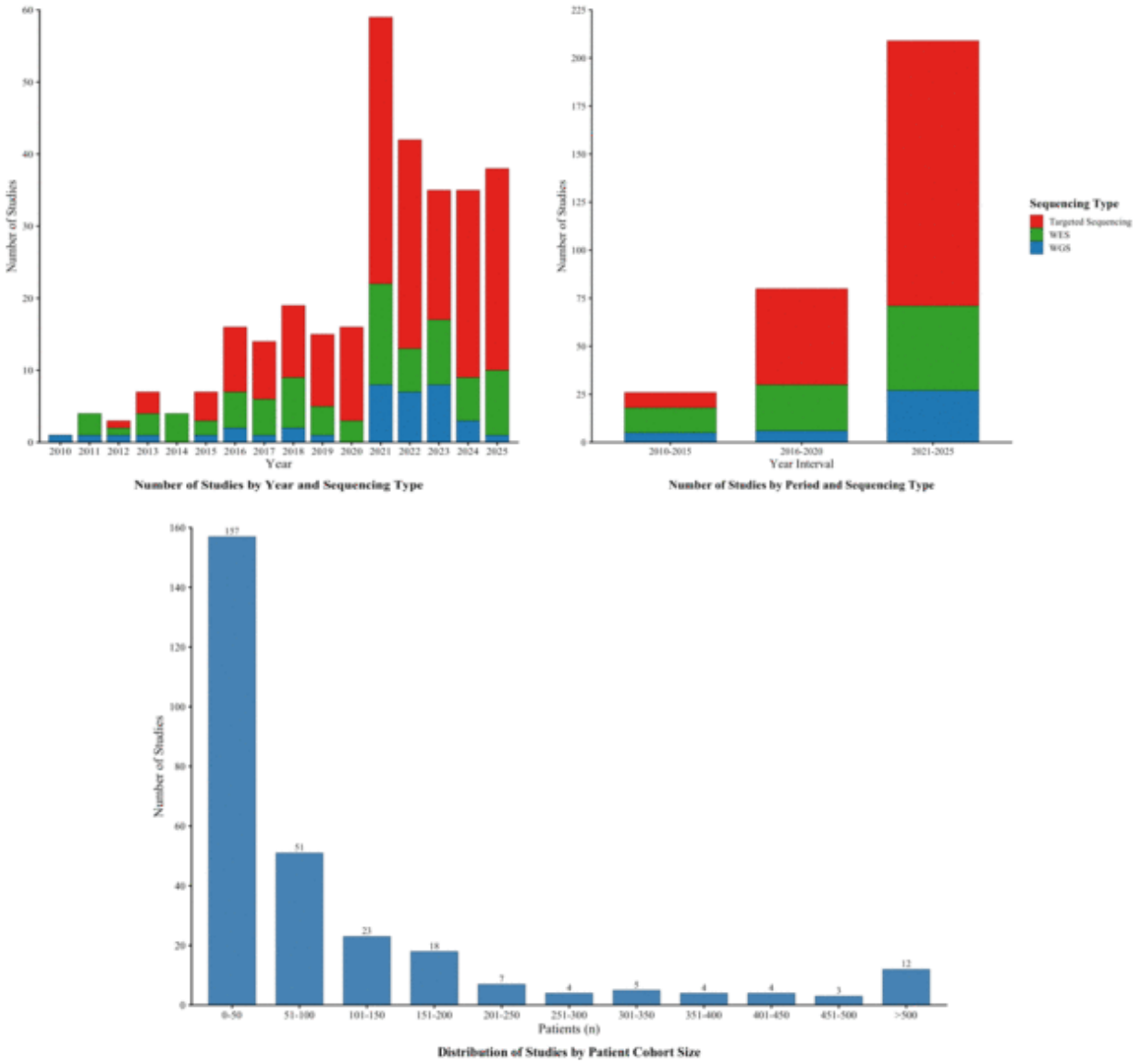

Figure 2. Characteristics and Trends of Included DLBCL Sequencing Studies
Stacked bar charts showing the annual and period-based counts of eligible studies stratified by sequencing type (targeted sequencing, WES, WGS), plus a histogram summarizing the distribution of cohort sizes. Highlighting that most studies were small, with substantial use of targeted sequencing across years.

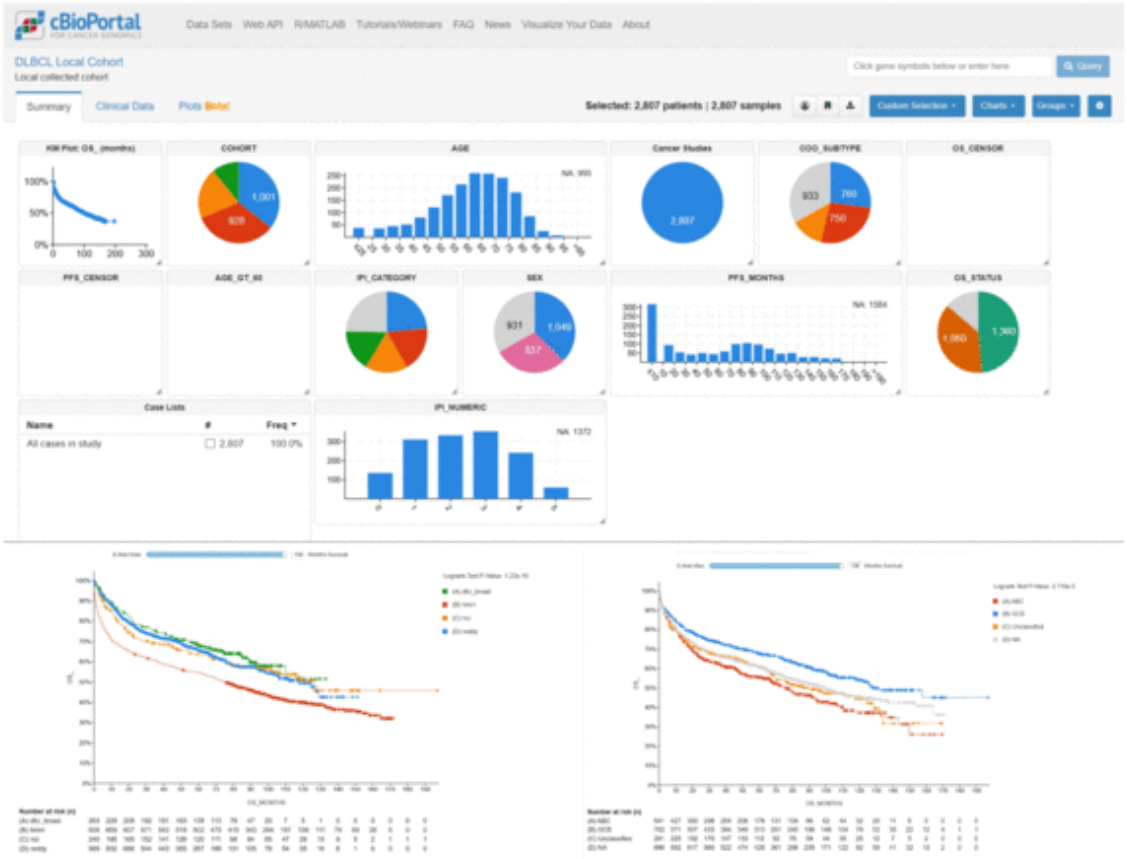

Figure 3. CLYDE interactive cBioPortal interface for cross-cohort clinico-genomic analysis
Dashboard view of the Clinico-Genomic Lymphoma Dataset Explorer (CLYDE), illustrating standardized core clinical variables, side-by-side cohort comparisons, subgroup survival analyses, and visualization of patient selection across datasets.