

Supplemental Materials:

A refined cell-of-origin classifier with targeted NGS and Artificial Intelligence shows robust predictive value in DLBCL

Supplemental Data File with two worksheets.xlsx

Supplemental Table S1. RNA quality control metrics

QC Metrics	Acceptance Criteria
RNA integrity	28s:18s rRNA ratio (Agilent 4200 tape station) (RIN): from 2 to 5
Total number of reads (limit of detection, LoD)	> 5 million reads
Base Quality	> 80% of bases with QS above > Q30
Percent Mapped to coding	> 60%
Mean Insert size	130-160 (bp)
Coverage	> 10 ×
% Cluster passing	The percent cluster passing filter (Cluster PF) > 80%
% Reads passing filter	The percent reads passing filter (Reads PF) > 80%
Positive Run Control	Mean expression level of the Universal Human Control =220±18 (Mean ± 2SD)
Limit of Blank (LoD)	< 1 ng at library preparation
% of duplicate reads	< 40%

Supplemental Table S2. Biomarkers showing differential distribution in the new GCB/ABC groups defined in the current study

Biomarker enriched in GCB or ABC	P value
Higher frequency in GCB vs ABC	
<i>BCL2</i> gene rearrangement	$P < 0.0001$
<i>MYC/BCL2</i> genetic double-hit	$P = 0.009$
BCL6+ expression	$P < 0.0001$
CD10+ expression	$P < 0.0001$
CTLA-4+ cell % in T cells	$P = 0.005$
GCET1+ expression	$P < 0.0001$
<i>PTEN</i> gene deletion	$P = 0.01$
CDKN2A p16+ expression	$P = 0.014$
SSBP2+ expression	$P < 0.0001$
Higher frequency in ABC vs GCB	
BCL2+ expression	$P = 0.002$
<i>BCL6</i> gene rearrangement	$P = 0.002$
CXCR4+ expression	$P = 0.014$
CD5+ expression	$P < 0.0001$
CD37+ expression	$P = 0.039$
FOXP1 high expression	$P < 0.0001$
IgM+ expression	$P < 0.0001$
JUNB high expression	$P = 0.002$
MUM1+ expression	$P < 0.0001$
MYC+/BCL2+ double expression	$P = 0.006$
NFKB1 p50+ expression	$P = 0.001$
CDKN1A/p21+ expression	$P = 0.003$
PD-L1 positive expression in B cells	$P < 0.0001$
High PD-L1+ cell % in B cells	$P = 0.007$
PD-L1 expression (by PD-L1+%) in CD4+ T cells	$P = 0.001$
PD-L1 expression (by PD-L1+%) in CD8+ T cells	$P < 0.0001$
PD-L1 expression (by PD-L1+%) in NK cells	$P = 0.001$
PRDM1+ expression	$P < 0.0001$
STAT3 high expression	$P = 0.001$

Supplemental Figure S1. A flow chart for the current study applying NGS and Artificial Intelligence (AI) in DLBCL model construction

