Supplementary material

Molecular classification of mature aggressive B cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens
Masque-Soler N¹, Szczepanowski M¹ Kohler C W², Spang R², Klapper W¹.

¹ Department of Pathology, Hematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel / Christian-Albrechts University, Kiel, Germany
² Institute of Functional Genomics, University of Regensburg, Germany

Supplementary Methods

RNA-extraction and assessment of quality

Formalin-fixed paraffin-embedded material was cut in five 10µm-thick pieces per sample. Extraction of RNA was done according to the manufacturer’s instructions (ExpressArt FFPE Clear RNAready Kit, AmpTec, Hamburg, Germany). Fresh-frozen blocks’ RNA was extracted as previously described ¹. To assess the RNA quality all 43 FFPE-extracted and the 2 newly extracted fresh-frozen RNA material was analyzed with the Agilent RNA 6000 Nano Chips (Agilent Technologies, Santa Clara, California, USA) following the product’s protocol. As expected, the quality of the FFPE samples were considerably degraded due to fixation steps but generally all samples yielded enough RNA with fragments at least longer than 200 nucleotides. A further material quality control is present after RNA processing at the nCounter instrument for digital multiplexed gene expression due to the company’s recommendations (see below for summary).

Digital multiplexed gene expression (DMGE)

Straight after a RNA quality measurement and with a known RNA concentration for each sample, 300ng were able to be hybridised with our custom-designed Reporter Probe (each with a specific 5’-end colour code signal) and Capture Probe (marked with biotin at the 3’ end). After the overnight hybridisation the samples were ready to load to the PrepStation where they were washed with a two-step magnetic bead-purification process and loaded in the 12-well cartridges. Afterwards the cartridges were loaded to the Digital Analyzer (DA, nCounter, NanoString Technologies Inc., Seattle, WA, USA). The DA resolution was set to high, meaning that 280 fields of view (FOV) were counted per sample. A comma-separated-value (.csv) file was obtained after processing. The first control on data-reliability was done with the FOV-counter to FOV-count ratio, which had to be >80%. Secondly, an acceptable binding density could not lay outside the 0,05-2,25 interval. Afterwards a data normalisation was done: individually adding up the internal positive controls and finding an average value in order to create a normalisation factor for each sample allowed for cartridge-to-cartridge comparison. The third quality control was visible at this point, since reliable samples had to show a normalisation factor between 0,3 and 3. Finally, through each sample’s average
internal negative control values, a background subtraction was performed. After these correction steps, nCounter data of FFPE and fresh frozen samples were normalized independently from each other using quantile normalization \(^2\) and finally log2 transformed afterwards.

**Study cohort**

All cases were previously characterized by the Molecular Mechanisms of Malignant Lymphoma (MMML\(^3\)) consortium using the Affymetrix GeneChip technology (referred to as the gold standard classification). We used the same original total RNA used for arrays for 48/50 (96%) samples (11 mBL, 12 intermediates and 27 non-mBL, the latter composed of 10 ABC, 8 GBC, 9 "unclassified"). For 2/50 (4%) cases the original RNA was exhausted. For these cases total RNA was newly extracted from the same fresh-frozen tissue blocks used previously for array based GEP. In addition, total RNA was isolated from corresponding FFPE tissue blocks in 40 cases (ExpressArt FFPE Clear RNAready Kit, AmpTec, Hamburg, Germany): 9 mBL, 8 intermediates, 23 non-mBL, the latter composed of 9 ABC, 8 GBC, 6 unclassified. For 39 cases GEPs from both fresh-frozen material and the FFPE tissue blocks were available (Supplementary Figure 1).

**Selection of genes and bioinformatics**

A subset of 5/74 probe sets from the original mBL signature (representing the genes SMARCA4, CD44, RNASEH2B, PRKAR2B, and CTSH) proved to be sufficient to classify mBL and non-mBL cases correctly. However, we also found that the small number of signature genes comes with a price. The robustness of the classification with respect to measurement errors is reduced. We thus selected one back-up gene for all five genes. This was always the gene on the microarray with the highest correlation to the signature gene across all training samples \(^1\). These 5 genes (TCF3, STAT3, NASP, PRDM10, BCL2A1) together with the 5 previous genes form the mBL set of the assay. In the same way, we chose 10 genes from the GCB/ABC signature \(^4\) and supplemented them with 10 back-up genes, giving us the ABC/GCB set of the assay. Supplementary table 1 gives a complete list of the classifier’s 30 genes with their NCBI reference numbers.

With our final 30 gene assay we obtained GEP for 50 fresh-frozen samples from mature aggressive B-cell lymphomas using DMGE. To obtain molecular classifications of “mBL”, “intermediate”, and “non-mBL” from DMGE data, we first trained a linear regression function to predict the mBL index \(^1\) of lymphomas using the 10 genes of the mBL set. We used least-angle regression \(^5\) and predicted all mBL-index values in leave-one-out cross-validation. We then calibrated two cut-off values C1 and C2 on the predicted mBL-index values and labeled a case “mBL” if the predicted index was below C1. If it was above C2 we labeled it “non-mBL”, and otherwise “intermediate”. This procedure was repeated independently for frozen and FFPE derived data.
Supplementary figure 1: Description of the study cohort
Supplementary figure 2: (A) Molecular classification of mature aggressive B-cell lymphomas as mBL, non-mBL and intermediates using kryo-block (fresh frozen)-derived RNA. The expression of the genes is displayed in a color code with high expression in yellow and low expression in blue. The molecular labels assigned by the array using fresh-frozen RNA and DMGE (nCounter) assay using fresh frozen-derived RNA are indicated as colored labels in the top bars. (B) Molecular classification of non-mBL using fresh frozen-derived RNA. The classification according to the immunohistochemical staining (Hans), the array using fresh-frozen RNA and DMGE assay using fresh frozen-derived RNA are indicated as colored labels in the top bars.
## Supplementary tables

### GCB/ABC signature

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>BATF</td>
<td>NM_006399.3</td>
</tr>
<tr>
<td>C3orf37</td>
<td>NM_020187.2</td>
</tr>
<tr>
<td>CCND2</td>
<td>NM_001759.3</td>
</tr>
<tr>
<td>CPNE3</td>
<td>NM_003909.2</td>
</tr>
<tr>
<td>DBI</td>
<td>NM_001079862.1</td>
</tr>
<tr>
<td>ENTPD1</td>
<td>NM_001098175.1</td>
</tr>
<tr>
<td>IGHM</td>
<td>ENST00000390559.1</td>
</tr>
<tr>
<td>IGLJ3</td>
<td>ENST00000390324.1</td>
</tr>
<tr>
<td>IL16</td>
<td>NM_004513.4</td>
</tr>
<tr>
<td>ITPKB</td>
<td>NM_002221.3</td>
</tr>
<tr>
<td>LRMP</td>
<td>NM_006152.2</td>
</tr>
<tr>
<td>MME</td>
<td>NM_00902.2</td>
</tr>
<tr>
<td>MYBL1</td>
<td>NM_001080416.2</td>
</tr>
<tr>
<td>PIM1</td>
<td>NM_002648.2</td>
</tr>
<tr>
<td>RAB7L1</td>
<td>NM_001135664.1</td>
</tr>
<tr>
<td>SH3BP5</td>
<td>NM_001018009.2</td>
</tr>
<tr>
<td>SKAP2</td>
<td>NM_003930.3</td>
</tr>
<tr>
<td>TBXA2R</td>
<td>NM_001060.3</td>
</tr>
<tr>
<td>TPD52</td>
<td>NM_001025252.1</td>
</tr>
</tbody>
</table>

### mBL signature

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCL2A1</td>
<td>NM_004049.2</td>
</tr>
<tr>
<td>CD44</td>
<td>NM_001001392.1</td>
</tr>
<tr>
<td>CTSH</td>
<td>NM_148979.2</td>
</tr>
<tr>
<td>NASP</td>
<td>NM_172164.1</td>
</tr>
<tr>
<td>PRDM10</td>
<td>NM_0020228.2</td>
</tr>
<tr>
<td>PRKAR2B</td>
<td>NM_002736.2</td>
</tr>
<tr>
<td>RNASEH2B</td>
<td>NM_001142279.1</td>
</tr>
<tr>
<td>SMARCA4</td>
<td>NM_003072.3</td>
</tr>
<tr>
<td>STAT3</td>
<td>NM_139276.2</td>
</tr>
<tr>
<td>TCF3</td>
<td>NM_003200.2</td>
</tr>
</tbody>
</table>

### endogenous controls

<table>
<thead>
<tr>
<th>Gene</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUSB</td>
<td>NM_000181.1</td>
</tr>
<tr>
<td>EPHB3</td>
<td>NM_004443.3</td>
</tr>
<tr>
<td>INE1</td>
<td>NR_024616.1</td>
</tr>
<tr>
<td>EML2</td>
<td>NM_012155.2</td>
</tr>
</tbody>
</table>

**Supplementary Table 1: Genes analyzed**
<table>
<thead>
<tr>
<th>GCB/ABC classifier</th>
<th>SAMPLE</th>
<th>IHS (Hans)</th>
<th>Array</th>
<th>DMGE frozen</th>
<th>DMGE FFPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPI-347</td>
<td>non-GCB</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
</tr>
<tr>
<td>MPI-407</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-584</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-619</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-626</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-698</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-700</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td><strong>MPI-795</strong></td>
<td>non-GCB</td>
<td>ABC</td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
<tr>
<td><strong>MPI-812</strong></td>
<td>N/A</td>
<td>ABC</td>
<td>ABC</td>
<td><strong>unclassified</strong></td>
<td></td>
</tr>
<tr>
<td>MPI-822</td>
<td>non-GCB</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
</tr>
<tr>
<td>MPI-831</td>
<td>N/A</td>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
<tr>
<td>MPI-841</td>
<td>N/A</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
</tr>
<tr>
<td><strong>MPI-842</strong></td>
<td>GCB</td>
<td><strong>ABC</strong></td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
<tr>
<td>MPI-846</td>
<td>non-GCB</td>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
<tr>
<td>MPI-847</td>
<td>non-GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td><strong>MPI-848</strong></td>
<td>N/A</td>
<td><strong>unclassified</strong></td>
<td>GCB</td>
<td><strong>GCB</strong></td>
<td></td>
</tr>
<tr>
<td>MPI-850</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-852</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-859</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-860</td>
<td>N/A</td>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
<tr>
<td>MPI-864</td>
<td>non-GCB</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
</tr>
<tr>
<td>MPI-865</td>
<td>N/A</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
</tr>
<tr>
<td>MPI-869</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-873</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-875</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-881</td>
<td>non-GCB</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
<td>ABC</td>
</tr>
<tr>
<td><strong>MPI-900</strong></td>
<td>GCB</td>
<td><strong>GCB</strong></td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
<tr>
<td>MPI-901</td>
<td>N/A</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
<td>GCB</td>
</tr>
<tr>
<td>MPI-902</td>
<td>non-GCB</td>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
<tr>
<td>MPI-905</td>
<td>N/A</td>
<td>unclassified</td>
<td>unclassified</td>
<td>unclassified</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>mBL classifier</th>
<th>GEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAMPLE</td>
<td>Array</td>
</tr>
<tr>
<td>MPI-043</td>
<td>mBL</td>
</tr>
<tr>
<td>MPI-084</td>
<td>mBL</td>
</tr>
<tr>
<td>MPI-347</td>
<td>non-mBL</td>
</tr>
<tr>
<td>MPI-407</td>
<td>intermediate</td>
</tr>
<tr>
<td>MPI-528</td>
<td>mBL</td>
</tr>
<tr>
<td>MPI-584</td>
<td>non-mBL</td>
</tr>
</tbody>
</table>
Supplementary Table 2: Complete list of classification according to array, IHS (Hans Algorithm 6), DMGE using kryo-derived RNA, DMGE using FFPE derived RNA. Lymphomas with divergent classification between the assays are labeled in red. Only lymphomas of the final cohort for which both array and DMGE data were available are shown. DMGE: digital multiplex gene expression.


