Supplemental methods

Next generation sequencing and filtering of variants
The Agilent HaloPlex target enrichment system was used to prepare DNA libraries from the diagnostic samples. The sequencing procedure was performed on a NextSeq500 (Illumina) for a custom panel of genes (all exons) involved in hematological malignancies. The number of genes included in the panel increased over time, and therefore not all the genes in the algorithm proposed by Gerstung et al.¹ was sequenced (see Supplemental Figure 2 and supplemental Table 2). After demultiplexing and generation of BAM files, variants were called and filtered with two commercial pipelines (SureCall, Agilent; and NextGENe, Soft Genetics). The mean coverage of the genes is reported in supplemental table 2. Variants of low frequency (Variant allele frequency < 2%) were excluded, as well as those with low depth sequencing (less than 12 reads of the mutant allele, or less than 30 total reads). We also excluded variants described as a polymorphism with a minor allele fraction (MAF) above 1% in dbSNP or ExAC and those with a MAF between 0.1% and 1% if variant allele frequency was between 40% and 60%.

Comparison of the prognosis scoring system performance
The probability of overall survival, non-remission death, non-relapse death, and relapse death were calculated using the R programs provided by Gerstung et al.¹ Missing clinical and molecular annotations were imputed following the approach proposed by Gerstung et al.¹

The ability of the different scores (knowledge databank approach, ELN 2010, ELN 2017) to discriminate patients who died during the first 3 years from those who did not die was assessed by receiver operating characteristic (ROC) curves, and the associated area under the curve. Due to censoring at 3 years, a time-dependent approach was used to calculate the area under the ROC curves. For non-remission death, non-relapse death, and relapse death analyses a competitive risk approach was used to build and calculate the area under the ROC curves.²

All the analyses were performed using the R software (R core Team, 2016), using the timeROC package described by Blanche et al.²


Supplemental Table 1: Survival status of the patients at 3 years after diagnosis

<table>
<thead>
<tr>
<th>Description</th>
<th>N = 155</th>
</tr>
</thead>
<tbody>
<tr>
<td>Censored</td>
<td>59 (38.1%)</td>
</tr>
<tr>
<td>Alive</td>
<td>31 (20.0%)</td>
</tr>
<tr>
<td>Non-relapse death</td>
<td>18 (11.6%)</td>
</tr>
<tr>
<td>Non-remission death</td>
<td>22 (14.2%)</td>
</tr>
<tr>
<td>Relapse death</td>
<td>25 (16.1%)</td>
</tr>
</tbody>
</table>

Supplemental figures legends:
**Supplemental Figure 1:** Kaplan-Meier estimation of overall survival probability for the 155 patients. The 95%CI are shown (discontinuous line)

**Supplemental Figure 2:** Heatmaps showing in each of the 155 patients the level of exhaustivity achieved for the clinical, cytogenetical, and molecular data (first 3 lanes), the mean follow-up, and outcome. Patients who did not undergo ASCT are presented in the top panel, whereas those who underwent ASCT are presented in the bottom panel.

**Supplemental Table 2:** detailed characteristics of the patients.

The clinical, cytogenetical and molecular data required for the algorithm are depicted for every patient.

For the molecular data, the mean sequencing depth of the genes is given in coma after the name of the gene.

The induction therapy regimen are grouped in three categories:

- **3+7 induction regimen:** intensive induction regimen based on anthracycline (3 days) and aracytine (7 days). Patients were included in the following clinical trials or treated accordingly (NCT00428558, NCT00880243, NCT00927498, NCT02473146, NCT01966497, NCT02416388)

- **Sequential induction regimen:** patients including in NCT00932412 clinical trial or treated accordingly

- **3+7+ATRA:** patients including in NCT00378365 clinical trial or treated accordingly

**Abbreviations:**

- nd: not determined
- sex: M: male, F: female;