Additional information and supplementary files

S1 **Ontologies that make up the lymphomyeloid cluster**. List of ontologies included in the lymphomyeloid cluster, used to devise the sensitiviy and specificity indicators in this manuscript.

S2 **Performance statistics of models trained in this study.** A table containing summary statistics of the performance of the models trained on the ontologies of the lymphomyeloid cluster and of non-lymphomyeloid ontologies is provided as supplementary data; the full list of results can be also be downloaded from the GitHub repository.

S3 **Performance map of model configurations based on other measures of predictive performance** While it based on accuracy throughout the paper, the sensitivity and specificity indicators can also be used to display the efficiency of each model configuration based on other metrics: F-score, precision and recall. The results are qualitatively identical.

S4 **Interactive plot of the performance map.** Code to run a Bokeh interactive plot of the statistical descriptor of accuracy for all ontologies, lymphomyeloid ontologies and non-lymphomyeloid ontologies. This can be run using `bokeh serve --show S4_performance_plot.py`.

S5 An interactive plot of the accuracy of the best model configuration (configuration PF:b using P sorting and the full feature set, FS11) projected on the an ontology embedding.

The authors declare that there are no competing interests regarding the publication of this research article.