SUPPLEMENTAL DATA

Gene expression-based model predicts outcome in children with intermediate-risk classical Hodgkin lymphoma

Johnston*, Mottok*, Chan, et al.

SUPPLEMENTAL METHODS

Patient cohorts and clinical-pathological characteristics

The training and independent validation cohorts for this study were drawn from patients enrolled in the Children's Oncology Group (COG) trial AHOD0031, a phase III clinical trial for newly diagnosed intermediate-risk cHL patients younger than 22 years of age. This trial was designed to combine dose-intensive chemotherapy with response-based treatment augmentation or reduction[1]. We selected 196 patients for the training cohort based on the availability of pre-treatment FFPET biopsies (paraffin blocks) from the Biopathology Center at the Cooperative Human Tissue Network. This cohort had clinical characteristics comparable to the overall AHOD0031 trial population (**Supplemental Table 4**). The validation cohort consisted of 84 patients for whom unstained tissue slides were available. This cohort was enriched for events (1:1 split), with events defined as relapse/progression, second malignant neoplasm, or death. Written informed consent was obtained according to institutional review board guidelines, and gene expression profiling (GEP) studies were approved by the BC Cancer/UBC review board (H12-01388, H19-00882).

Gene expression analysis

NanoString CodeSets: Total RNA extraction and NanoString digital GEP were performed as previously described using FFPET sections[2]. The study herein included two NanoString custom CodeSets: **PHL800** and **PHL-9C** (**Supplemental Tables 1 and 2**). Note PHL-9C refers to both the NanoString custom CodeSet and the prognostic model derived from the CodeSet. GEP of specimens from the training cohort were obtained using PHL800 containing 813 probes, including probes for 784 endogenous genes, 15 housekeeping genes, 8 negative controls, and 6 positive controls. The endogenous genes were selected based on suggested HL prognostic genes from the literature, and markers representative of HRS cells and TME components[2]. GEP of the validation cohort specimens were interrogated using PHL-9C, a 136-gene CodeSet containing probes for 111 endogenous genes (representing 9 cellular components), 11 housekeeping genes, 8 negative controls, and 6 positive controls. The gene list was based on the results of the feature selection process.

Normalization and quality control: All downstream gene expression analysis methods, including model building and statistical analyses, were performed using R version 3.1.1, and unless otherwise specified figures were produced using ggplot2 version 2.0.1. Normalization of gene expression values and quality control were performed using similar procedures described in Scott et al.[3] and Chan et al.[2] Specifically, normalization was performed by dividing the raw NanoString counts of each case by its normalizer score (calculated using the geometric

mean of 12 housekeeping genes: *ACTB*, *ALAS1*, *CLTC*, *GAPDH*, *GUSB*, *PGK1*, *POLR2A*, *RPL19*, *RPLP0*, *SDHA*, *TBP*, and *TUBB*), and then multiplying by 1000. The three remaining housekeeping genes included in the PHL800 CodeSet (*HMBS*, *POLR1B*, and *G6PD*) were removed from further analyses, as *HMBS* and *POLR1B* obtained the lowest median expression values across the samples, and *G6PD* is known to be X-linked.

Cases with very low normalizer scores often obtain very high normalized gene expression values, which indicates a poor-quality sample. To perform sample quality control, normalized gene expression of each of the 12 housekeeping genes was plotted against the normalizer score for each sample. Any specimen with expression greater than ±2 standard deviations (SD) of the mean of any of the 12 housekeeping genes was considered an outlier. To determine the threshold for classifying specimens as outliers, the mean normalized expression of all genes, excluding the 12 housekeeping genes and EBV related genes (*EBER1*, *EBER2*, *LMP1*, and *LMP2*) per case was plotted, and the outlier sample with the maximum normalizer score was chosen as the threshold to classify outlier samples. The threshold was selected to maximize the number of excluded specimens with abnormal housekeeping gene expression, while minimizing the number of excluded cases with expression within ±2 SD of the mean of each housekeeping gene. Using this approach, we determined the threshold for the normalizer score to be 23.82, and samples with normalizer scores less than the threshold were excluded from further analyses. Subsequently, normalized count data from specimens that passed quality control (n=175 and n=71 for the training and validation cohorts, respectively) were $\log_2$ transformed, and only endogenous genes were considered for further analyses.

Statistical analyses

Event free survival (EFS) was measured from time of study enrollment to event (relapse/progression, second malignant neoplasm, or death). Univariate Cox regression analysis was performed using the coxph and Surv functions from the R survival package version 2.40-1. To test for an interaction between gene expression (individual genes within the 23-gene adult cHL prognostic model) and study cohort (pediatric/adult), we used Cox regression analysis with an interaction term. Specifically, we used coxph(Surv(time, event) ~ GeneExpr*Study, which tests three effects: the main effect of a gene, the main effect of the cohort, and the interaction between a gene and cohort. Time and event were collapsed across the two cohorts (i.e. event-free survival was used for the pediatric cohort, and was collapsed with failure-free survival or overall survival for the adult cohort) and Study was coded as a factor (“1” for pediatric cohort and “0” for adult cohort). We then filtered the coxph results to the interaction (i.e. GeneExpr:Study) after adjusting for main effects of a gene and cohort. The false discovery rate (FDR) was determined using the p.adjust function from the stats package.

Survival curves were created with the survfit function from the survival package using the Kaplan Meier formula and the ggsurvplot function from survminer version 0.2.1. All survival curves were calculated using intention to treat analysis.

Cellular components

To de-convolute the cellular composition of tumors, we used gene expression signatures termed cellular component scores. A cellular component is defined as a cell type within the microenvironment of HL (including HRS cells), and a cellular component score is the median expression level of two or more genes per patient known to be associated with the cell type. Genes were assigned to various cellular components/signatures as per scientific literature, where a gene can belong to more than one signature (**Supplemental Tables 1 and 2**).

Comparing pediatric to adult cHL

To compare expression profiles between pediatric and adult cHL, we utilized gene expression data for 218 genes common to our training cohort (n=175) and a previously published adult cHL cohort (n=290)[3] (**Supplemental Table 1**). We removed technical bias by performing quantile normalization across batches, and by utilizing cellular component scores instead of the expression of singular genes. Raw expression values for genes common across the two cohorts were quantile normalized. Quantile normalization was performed using the function normalize.quantiles from the Bioconductor package preprocessCore version 1.28.0. Post-normalization, the probes *ACTB*, *B2M*, *CD74*, *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRA*, and *HLA-DRB3*, were removed from downstream analyses due to quantile normalization artifacts. The Spearman correlation of age and cellular component scores was calculated using the function cor.

Model building

We utilized a Cox regression model on gene expression data from 175 patients in our training cohort to build a model for EFS in pediatric cHL. The cv.glmnet function from the package glmnet (version 2.0-5) was used to fit the penalized Cox regression model. The lambda regularization parameter was trained by using a leave-one-out cross-validation approach. The optimal lambda value was chosen based on the minimum partial likelihood deviance. Input features to the penalized Cox regression model were cellular component scores and the alpha parameter was set to 0 for ridge regression. All genes per cellular component were used to calculate cellular component scores for model input, regardless of whether their expression was significantly associated with EFS.

The threshold for patient scores from the EFS prognostic model that separates patients into ‘low’ and ‘high’ risk groups was determined using the survdiff function from the survival

package. Values were tested in increments of 0.01, and the score that obtained the largest chi-square statistic from the log-rank test between the two risk groups was selected as the threshold. Time-dependent receiver operating characteristic (ROC) curves were produced using the package timeROC version 0.3 calculated at 0.1 year intervals.

To account for CodeSet variability and the necessity for threshold adjustment, 21 calibrator specimens from the training cohort were applied to the PHL-9C CodeSet. The adjustment was determined by correlating the scores obtained by the HL800 CodeSet (x-values) to the scores obtained by the PHL-9C CodeSet (y-values) for the 21 calibrator samples for each respective predictor model. A line of best fit was applied, and the y-value at the respective model risk score obtained by the HL800 CodeSet represented the adjusted risk threshold.

Log-rank tests and Cox proportional hazard models were implemented to test the prognostic ability of the PHL-9C model when used alone and in combination with other clinical factors using the coxph function from the survival package. Since our validation cohort was enriched for patients with events (1:1), a weighted analysis approach was also implemented to achieve an unbiased estimate of relative risk according to the weighted analysis methods as published[3] and originally proposed by Gray[4]. The non-event to event ratio was 5.58 in the entire AHOD0031 cohort and 1.22 in the validation cohort. Therefore, weights of 0.3438 and 1.5385 were assigned to cases with and without events in the validation cohort, respectively. The survey package 3.30.3 was used to perform the weighted log-rank test using the svylogrank function and implement a weighted Cox proportional hazards regression model using the svycoxph function.

Immunohistochemistry (IHC) and Epstein-Barr virus-encoded RNA in situ hybridization (EBER-ISH)

For the pediatric HL cohort, 155 samples were subjected to EBER-ISH, and classified as EBV-negative, EBV-positive HRS cells, or EBV background. For the remaining samples, the threshold for EBV positivity was determined by the minimum normalized gene expression value of EBER1 obtained by EBV-positive HRS cells using EBER-ISH. For the adult HL cohort, EBV status was determined using EBER-ISH. IHC for Thymus and activation-regulated chemokine (TARC) was performed and scored using a Ventana Benchmark system as published[5]. We used the polyclonal goat anti-TARC antibody (AF364, R&D systems, concentration 1:800). Antigen retrieval was performed with 10 mM citrate buffer at pH 6.0 using standard protocols with DAB visualization. Scoring was performed by assessing the percentage of positively stained tumor cells multiplied by staining intensity (0=negative, 1=weak, 2=moderate, 3=strong). Cases with a histoscore of less or equal to 100 were considered to be TARC low.

SUPPLEMENTAL TABLES

Supplemental Table 1: PHL800 NanoString custom CodeSet.

Attached as separate Excel file.

Supplemental Table 2: PHL-9C NanoString custom CodeSet.

Attached as separate Excel file.

Supplemental Table 3: Clinico-pathological characteristics of the study cohorts.

Characteristic	Training cohort (n = 175)	Validation cohort (n = 71)	*P*
Age, years			
Median (range)	15 (1-21)	15 (3-21)	.53
Male, %	51	48	.67
Ann Arbor stage, %			.75
I	5	6	
II	58	56	
III	21	17	
IV	16	21	
Histological subtype, No. (%)			.14
Nodular sclerosis	162 (93)	60 (85)	
Mixed cellularity	9 (5)	7 (10)	
Missing	4 (2)	4 (5)	
EBV-positive cases, No. (%)	39 (22)	nd	
Therapy response			
PET, No. (%)*			.14
Positive	27 (22)	17 (37)	
Negative	87 (71)	26 (56.5)	
Equivocal	8 (7)	3 (6.5)	
Stratum SER, No. (%)	30 (17)	22 (31)	.05

Abbreviations: EBV, Epstein-Barr virus; nd, not determined; SER, Slow Early Responder.
*PET examination was unavailable for all patients.

Supplemental Table 4: Clinico-pathological characteristics of patients in the training and validation cohort compared to remaining patients enrolled in AHOD0031.

	Training vs. Rest of AHOD0031			**Validation vs. Rest of AHOD0031**		
Characteristic	Training cohort (n = 175)	Rest of AHOD0031 (n = 1537)	*P*	Validation cohort (n = 71)	Rest of AHOD0031 (n = 1641)	*P*
Age, years			.31			.92
Mean	14.81	14.54		14.53	14.57	
Male, No. (%)	90 (51)	818 (53)	.69	34 (48)	874 (53)	.40
Stage, No. (%)			.96			.47
I	9 (5)	89 (6)		4 (6)	94 (6)	
II	102 (58)	904 (59)		40 (56)	966 (59)	
III	36 (21)	318 (21)		12 (17)	342 (21)	
IV	28(16)	226 (15)		15 (21)	239 (15)	
Histology, No. (%)			<.001			.16
Nodular sclerosis	162 (93)	1221 (79.4)		60 (85)	1323 (80.6)	
Mixed cellularity	9 (5)	147 (9.6)		7 (10)	149 (9.1)	
Lymphocyte predominant	0 (0)	97 (6.3)		0 (0)	97 (6.0)	
Lymphocyte depleted	0 (0)	4 (0.2)		0 (0)	4 (0.2)	
Missing/Unknown	4 (2)	68 (4.4)		4 (5)	68 (4.1)	
Therapy response						
PET, No. (%)*			.23			.08
Positive	27 (22)	255 (25)		17 (37)	265 (24)	
Negative	87 (71)	722 (71)		26 (56.5)	783 (72)	
Equivocal	8 (7)	36 (4)		3 (6.5)	41 (4)	
Stratum, No. (%)†			.98			.01
RER	142 (81)	1327 (81)		49 (69)	1320 (81)	
SER	30 (17)	275 (17)		22 (31)	283 (17)	
Missing/Unknown	3 (2)	32 (2)		0 (0)	35 (2)	

Abbreviations: RER, Rapid Early Responder; SER, Slow Early Responder.

*PET examination was unavailable for all patients in AHOD0031.

†Stratum was unavailable for all patients in AHOD0031.

Supplemental Table 5: Cox regression on each of the 23 genes in the adult cHL prognostic model[3] using EFS as outcome for the pediatric cohort and OS as outcome for the adult cohort[*].

	Pediatric cohort (n = 175)		Adult cohort (n = 290)		Interaction[†]	
Gene	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
CCL17	1.22 (1.02 - 1.45)	.025	0.85 (0.77 - 0.95)	.003	1.43 (1.17 - 1.75)	<.001
PDGFRA	1.19 (0.91 - 1.55)	.201	0.58 (0.42 - 0.81)	.001	2.02 (1.32 - 3.09)	.001
COL6A1	1.00 (0.78 - 1.29)	.996	0.59 (0.44 - 0.80)	<.001	1.65 (1.12 - 2.44)	.012
TNFSF10	0.96 (0.68 - 1.36)	.809	2.36 (1.51 - 3.70)	<.001	0.42 (0.24 - 0.73)	.002
RNF144B	0.93 (0.61 - 1.40)	.713	1.84 (1.09 - 3.10)	.021	0.51 (0.27 - 0.99)	.046
HLA-C	0.92 (0.54 - 1.58)	.767	2.00 (1.05 - 3.82)	.035	0.47 (0.20 - 1.10)	.081
CD300A	0.90 (0.66 - 1.24)	.523	2.30 (1.31 - 4.04)	.004	0.40 (0.21 - 0.76)	.005
CD68	0.90 (0.62 - 1.31)	.573	2.08 (1.31 - 3.28)	.002	0.43 (0.24 - 0.78)	.006
LYZ	0.81 (0.65 - 1.01)	.057	1.60 (1.12 - 2.30)	.011	0.51 (0.33 - 0.78)	.002
STAT1	0.78 (0.56 - 1.10)	.155	1.75 (1.13 - 2.72)	.012	0.45 (0.26 - 0.78)	.004
GLUL	0.78 (0.53 - 1.15)	.213	2.16 (1.33 - 3.53)	.002	0.36 (0.19 - 0.67)	.001
APOL6	0.77 (0.56 - 1.06)	.109	2.17 (1.10 - 4.31)	.026	0.35 (0.16 - 0.74)	.006
HLA-A	0.77 (0.48 - 1.23)	.267	1.88 (0.90 - 3.89)	.091	0.41 (0.17 - 0.99)	.046
CXCL11	0.74 (0.60 - 0.91)	.004	1.46 (1.20 - 1.79)	<.001	0.50 (0.38 - 0.67)	<.001
WDR83	0.73 (0.46 - 1.14)	.170	1.68 (0.72 - 3.90)	.227	0.43 (0.17 - 1.12)	.084
ALDH1A1	0.73 (0.56 - 0.94)	.014	1.59 (1.22 - 2.08)	<.001	0.46 (0.32 - 0.67)	<.001
IRF1	0.72 (0.47 - 1.10)	.128	2.15 (1.28 - 3.63)	.004	0.34 (0.17 - 0.66)	.002
IL15RA	0.67 (0.40 - 1.11)	.117	2.05 (1.03 - 4.10)	.041	0.32 (0.14 - 0.76)	.010
LMO2	0.63 (0.44 - 0.89)	.010	2.41 (1.31 - 4.44)	.005	0.26 (0.13 - 0.53)	<.001
B2M	0.60 (0.25 - 1.41)	.239	1.74 (0.21 - 14.37)	.605	0.34 (0.04 - 3.36)	.359
IFNG	0.58 (0.43 - 0.78)	<.001	1.57 (1.20 - 2.07)	.001	0.37 (0.25 - 0.55)	<.001
PRF1	0.53 (0.34 - 0.83)	.005	1.72 (1.20 - 2.46)	.003	0.30 (0.17 - 0.53)	<.001
RAPGEF2	0.51 (0.23 - 1.12)	.092	2.28 (1.04 - 4.97)	.039	0.23 (0.07 - 0.68)	.008

Abbreviations: CI, Confidence Interval; HR, Hazard Ratio.

[*]Values for pediatric and adult cohorts were used to create Supplemental Figures 2F and 2G, respectively. Genes ordered according to descending HR for the pediatric cohort.

[†]Test for interaction between gene expression and cohort (pediatric/adult). $P < 0.05$ indicates that the effect of the gene is not consistent between the pediatric and adult cohorts.

Supplemental Table 6 Cox regression on each of the 23 genes in the adult cHL prognostic model[6] using EFS as outcome for pediatric cohort and failure-free survival as outcome for adult cohort.

	Pediatric cohort (n = 175)		Adult cohort (n = 290)		Interaction*	
Gene	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
CCL17	1.22 (1.02 - 1.45)	.025	0.91 (0.85 - 0.99)	.027	1.33 (1.10 - 1.61)	.003
PDGFRA	1.19 (0.91 - 1.55)	.201	0.80 (0.65 - 1.00)	.045	1.48 (1.05 - 2.09)	.025
COL6A1	1.00 (0.78 - 1.29)	.996	0.87 (0.72 - 1.05)	.153	1.15 (0.84 - 1.58)	.384
TNFSF10	0.96 (0.68 - 1.36)	.809	1.34 (0.96 - 1.88)	.085	0.71 (0.44 - 1.15)	.164
RNF144B	0.93 (0.61 - 1.40)	.713	1.71 (1.20 - 2.44)	.003	0.54 (0.31 - 0.92)	.025
HLA-C	0.92 (0.54 - 1.58)	.767	1.01 (0.67 - 1.53)	.949	0.91 (0.46 - 1.80)	.784
CD300A	0.90 (0.66 - 1.24)	.523	1.41 (0.97 - 2.06)	.075	0.64 (0.39 - 1.05)	.077
CD68	0.90 (0.62 - 1.31)	.573	1.54 (1.13 - 2.11)	.007	0.58 (0.35 - 0.94)	.028
LYZ	0.81 (0.65 - 1.01)	.057	1.40 (1.09 - 1.80)	.009	0.57 (0.41 - 0.80)	.001
STAT1	0.78 (0.56 - 1.10)	.155	1.29 (0.98 - 1.70)	.064	0.60 (0.39 - 0.93)	.022
GLUL	0.78 (0.53 - 1.15)	.213	1.47 (1.06 - 2.03)	.021	0.53 (0.32 - 0.88)	.014
APOL6	0.77 (0.56 - 1.06)	.109	1.36 (0.86 - 2.17)	.190	0.56 (0.32 - 0.99)	.046
HLA-A	0.77 (0.48 - 1.23)	.267	1.21 (0.75 - 1.96)	.429	0.63 (0.32 - 1.23)	.176
CXCL11	0.74 (0.60 - 0.91)	.004	1.20 (1.05 - 1.37)	.006	0.61 (0.48 - 0.78)	<.001
WDR83	0.73 (0.46 - 1.14)	.17	1.32 (0.74 - 2.33)	.345	0.55 (0.27 - 1.14)	.110
ALDH1A1	0.73 (0.56 - 0.94)	.014	1.32 (1.09 - 1.59)	.004	0.55 (0.40 - 0.75)	<.001
IRF1	0.72 (0.47 - 1.10)	.128	1.46 (1.02 - 2.09)	.039	0.49 (0.28 - 0.85)	.012
IL15RA	0.67 (0.40 - 1.11)	.117	1.96 (1.22 - 3.16)	.006	0.33 (0.17 - 0.67)	.002
LMO2	0.63 (0.44 - 0.89)	.010	1.46 (0.98 - 2.17)	.060	0.43 (0.25 - 0.73)	.002
B2M	0.60 (0.25 - 1.41)	.239	3.34 (0.60 - 18.58)	.169	0.18 (0.03 - 1.21)	.078
IFNG	0.58 (0.43 - 0.78)	<.001	1.34 (1.11 - 1.61)	.002	0.43 (0.31 - 0.61)	<.001
PRF1	0.53 (0.34 - 0.83)	.005	1.54 (1.20 - 1.98)	<.001	0.35 (0.21 - 0.57)	<.001
RAPGEF2	0.51 (0.23 - 1.12)	.092	1.34 (0.77 - 2.34)	.300	0.38 (0.14 - 0.98)	.046

Abbreviations: CI, Confidence Interval; HR, Hazard Ratio.
*Test for interaction between gene expression and cohort (pediatric/adult). $P < 0.05$ indicates that the effect of the gene is not consistent between the pediatric and adult cohorts.

Supplemental Table 7: The 9-cellular component EFS model cellular components and coefficients.

Component	Coefficient
B-cell	-0.1792
CTL	-0.2942
FDC	0.2763
HRS cell	0.0584
Mast cell	0.1850
MDSC	0.3528
Th1 cell	-0.1376
Th2 cell	0.1142
Treg cell	0.2684

Supplemental Table 8: Univariate and multivariable Cox regression analyses in the pediatric HL validation cohort.

Variable	Patients		Univariate Cox regression			Multivariable Cox regression*		
	No.	%	HR	SE	*P*	HR	SE	*P*
Model score high	31	43.7	2.49	0.36	.012	2.81	0.40	.009
Stage IV	15	21.1	1.58	0.39	.246			
Mediastinal mass > 0.33†	25	35.7	1.88	0.35	.076			
Fever	13	18.3	2.79	0.40	.009	3.19	0.38	.003
Albumin < 3.5‡	19	28.4	2.66	0.38	.010	2.08	0.40	.068
Stratum SER	22	31.0	1.15	0.37	.714			

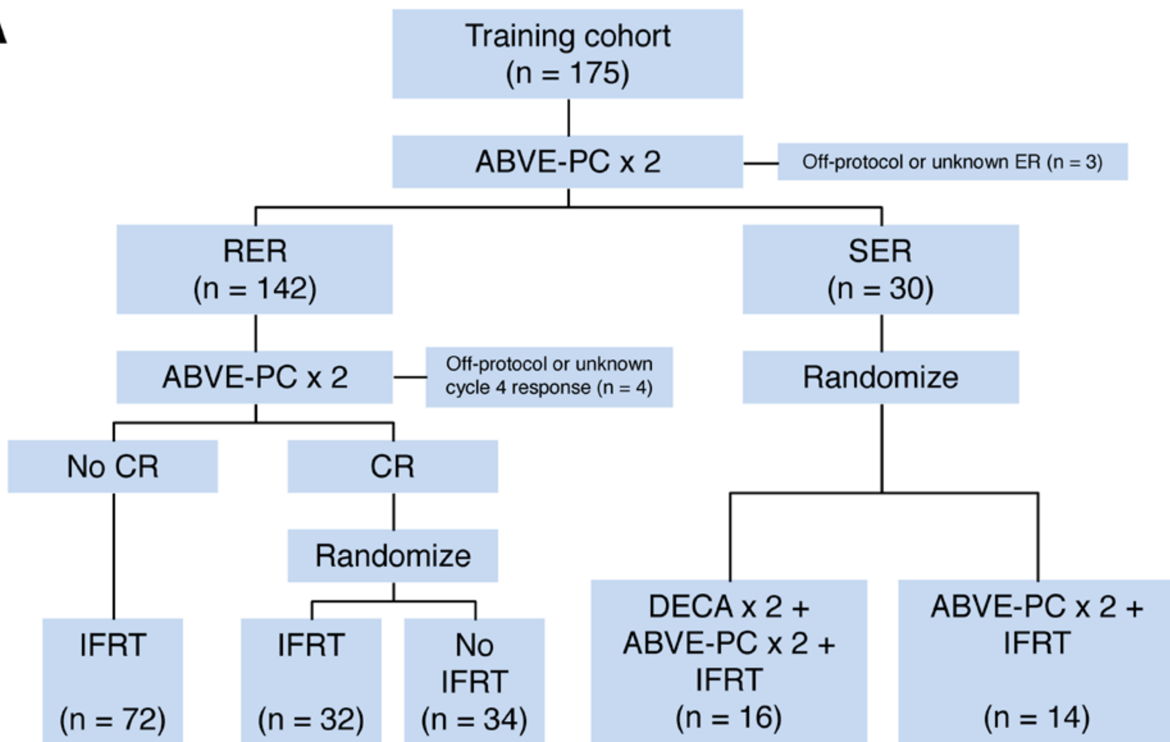
Abbreviations: HR, Hazard Ratio; SE, Standard Error; SER, Slow Early Responder.

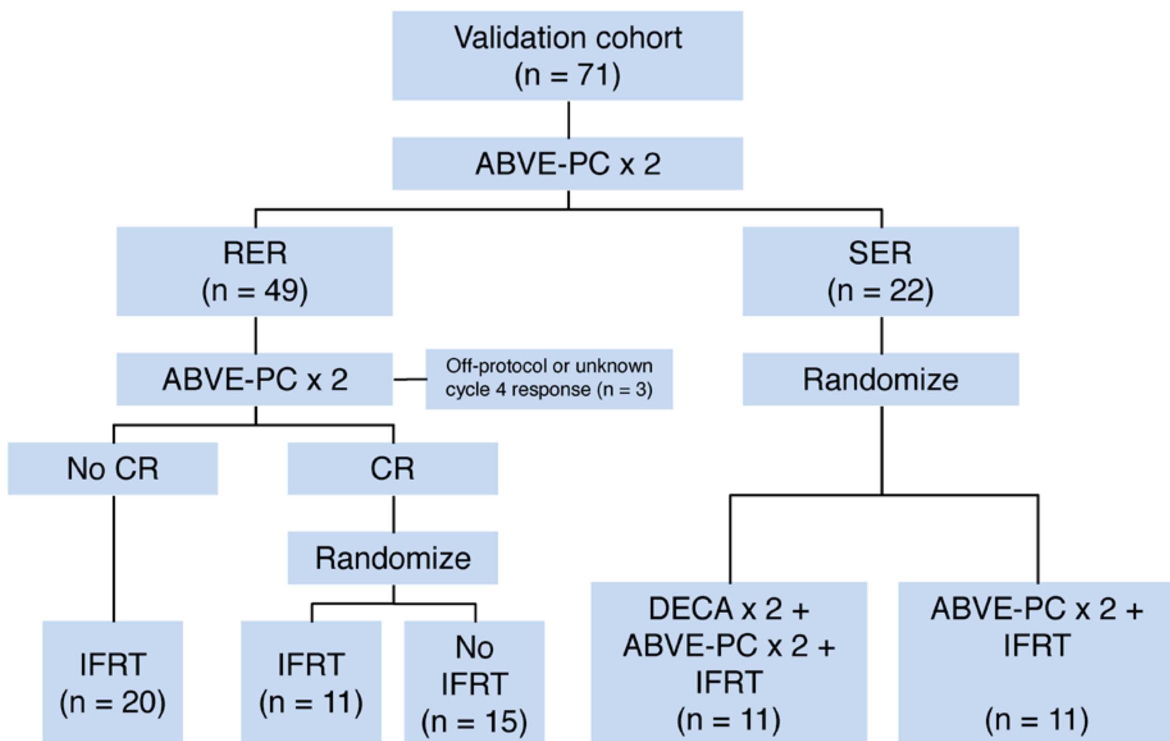
*Variables with $P < .05$ were entered into the multivariable Cox model.

†Mediastinal mass was unavailable for one patient.

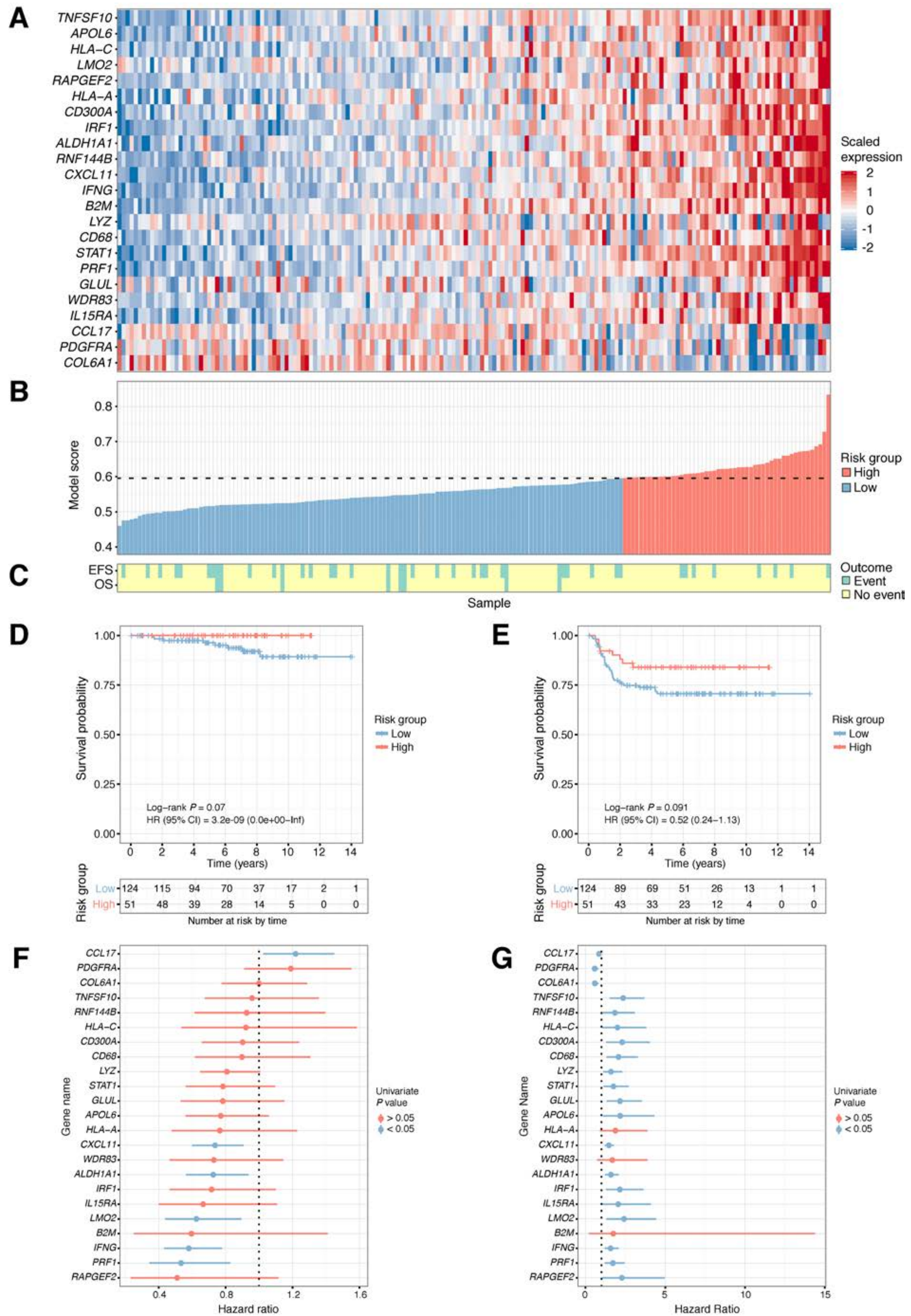
‡Albumin measurements were unavailable for four patients.

SUPPLEMENTAL FIGURES

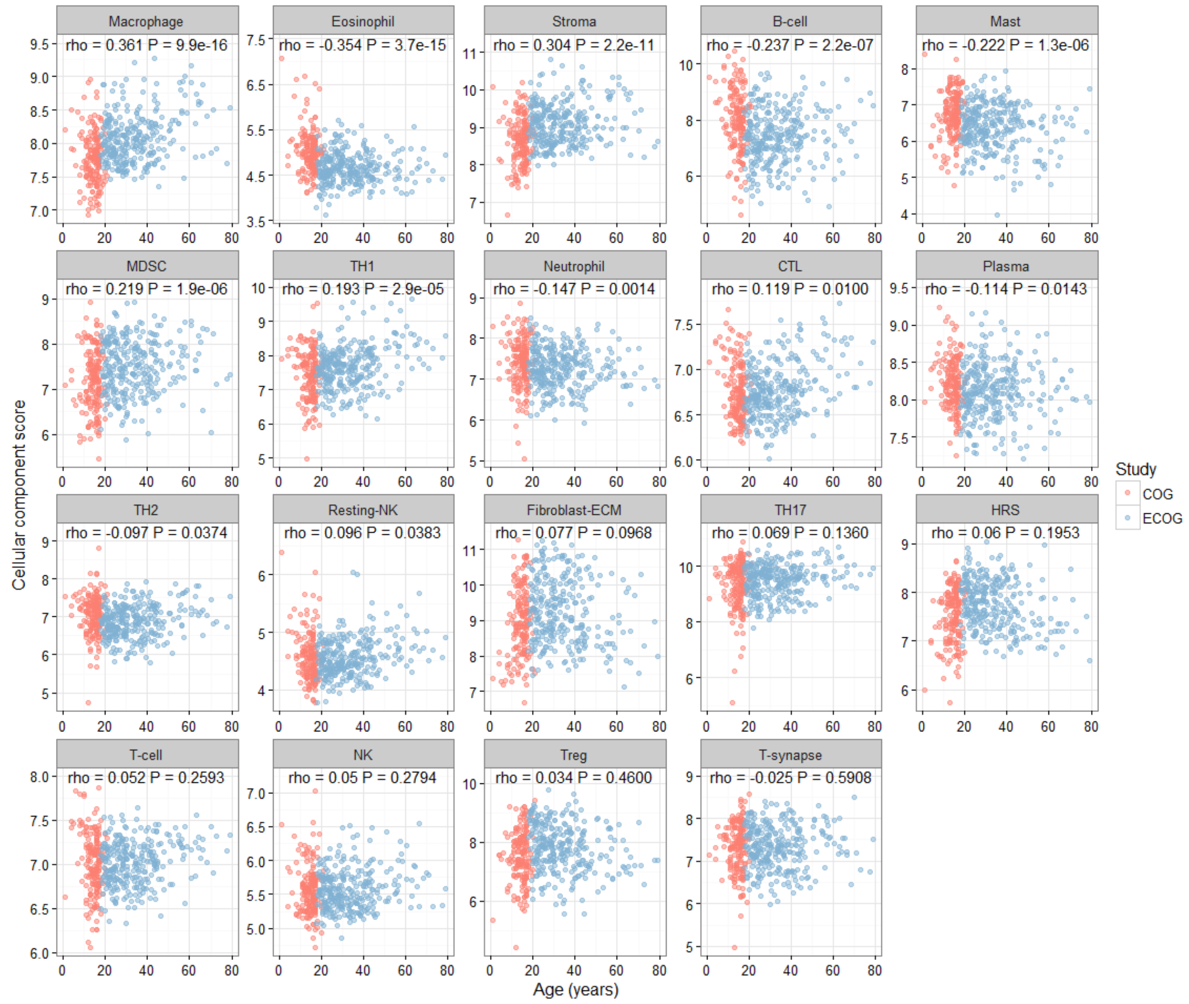




Supplemental Figure 1: Distribution of treatment regimens in the training and validation cohorts according to the AHOD0031 experimental design schema. Distribution of patients per treatment regimen for the training cohort **(A)** and validation cohort **(B).** ABVE-PC: Doxorubicin, Bleomycin, Vincristine, Etoposide, Prednisone, Cyclophosphamide; CR: Complete Response; CT: Computed Tomography; DECA: Dexamethasone, Etoposide, Cisplatin, Cytarabine; ER: Early Response; IFRT: Involved-Field Radiation Therapy; RER: Rapid Early Responder; SER: Slow Early Responder.

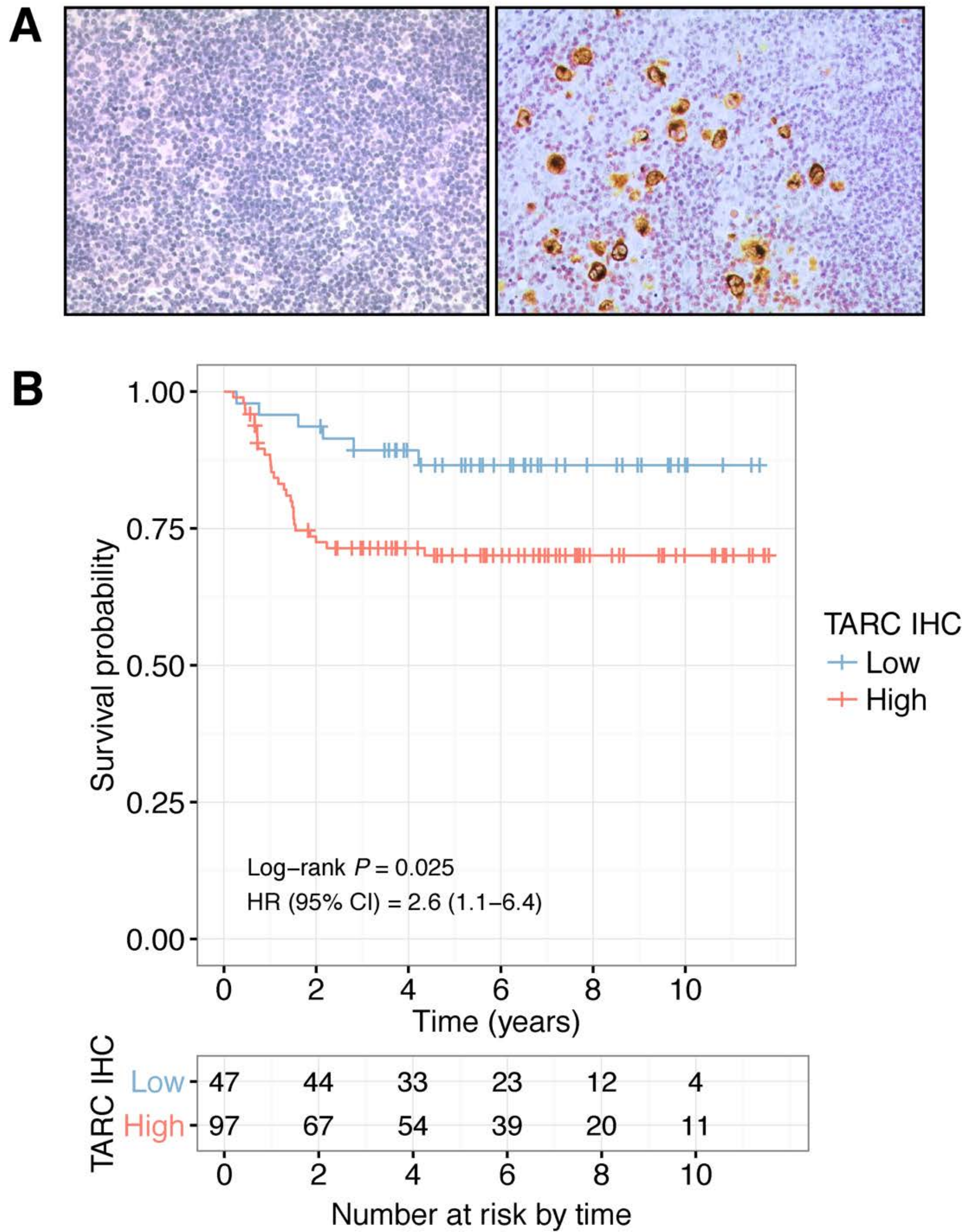

A
TNFSF10
APOL6
HLA-C
LMO2
RAPGEF2
HLA-A
CD300A
IRF1
ALDH1A1
RNF144B
CXCL11
IFNG
B2M
LYZ
CD68
STAT1
PRF1
GLUL
WDR83
IL15RA
CCL17
PDGFRA
COL6A1
Scaled expression
B
Model score
Risk group
High
Low
C
EFS
OS
Sample
Outcome
Event
No event
D
Survival probability
Log-rank P = 0.07
HR (95% CI) = 3.2e-09 (0.0e+00-Inf)
Time (years)
Risk group
Low
High
Number at risk by time
E
Log-rank P = 0.091
HR (95% CI) = 0.52 (0.24-1.13)
F
Gene name
Hazard ratio
Univariate P value
> 0.05
< 0.05
G
Gene Name
Hazard Ratio

Supplemental Figure 2: The 23-gene outcome predictor for adult cHL applied to the pediatric cHL training cohort. (A) Scaled gene expression values of the 23 genes in the prognostic model for adult cHL developed by Scott et al. Columns represent patients arranged by model score, and rows represent genes arranged by model coefficient. **(B)** Model scores for the 23-gene model and assigned risk classes as defined by the model score threshold (dotted line). **(C)** Survival outcomes of the patients in the training cohort. Kaplan Meier estimates of OS **(D)** and EFS **(E)**, where low- and high-risk were determined by 23-gene predictor scores as per (B). **(F)** Univariate Cox regression of EFS and each of the 23 genes in the adult cHL prognostic model for patients in the pediatric cHL training cohort. **(G)** Univariate Cox regression of OS and each of the 23 genes in the adult cHL prognostic model for patients in the adult cHL cohort.

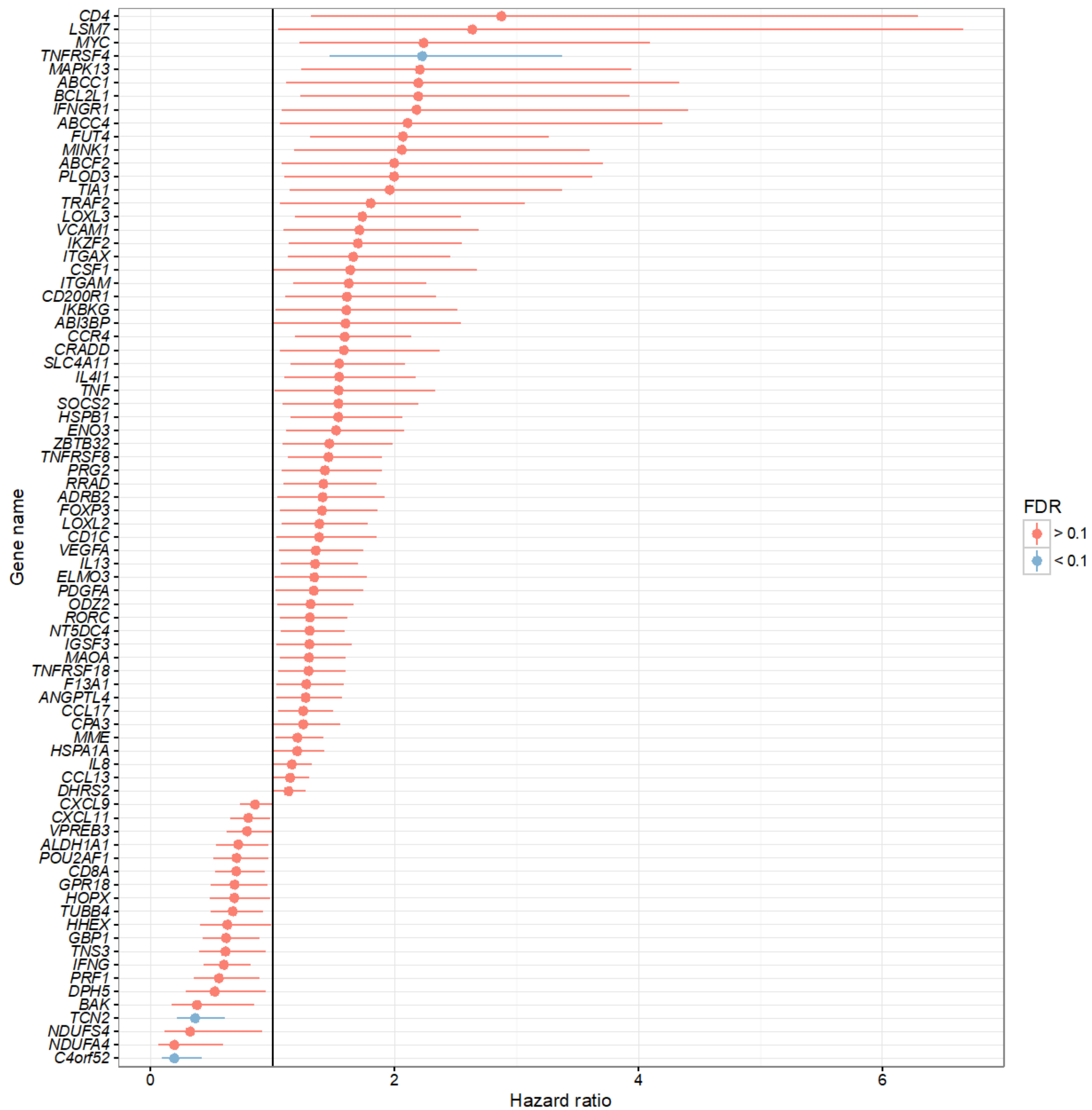


Supplemental Figure 3: Spearman correlation of patient age and cellular component score for all patients across pediatric and adult HL cohorts. Cellular components ordered by Spearman correlation *P*. Red points: pediatric cohort; blue points: adult cohort.

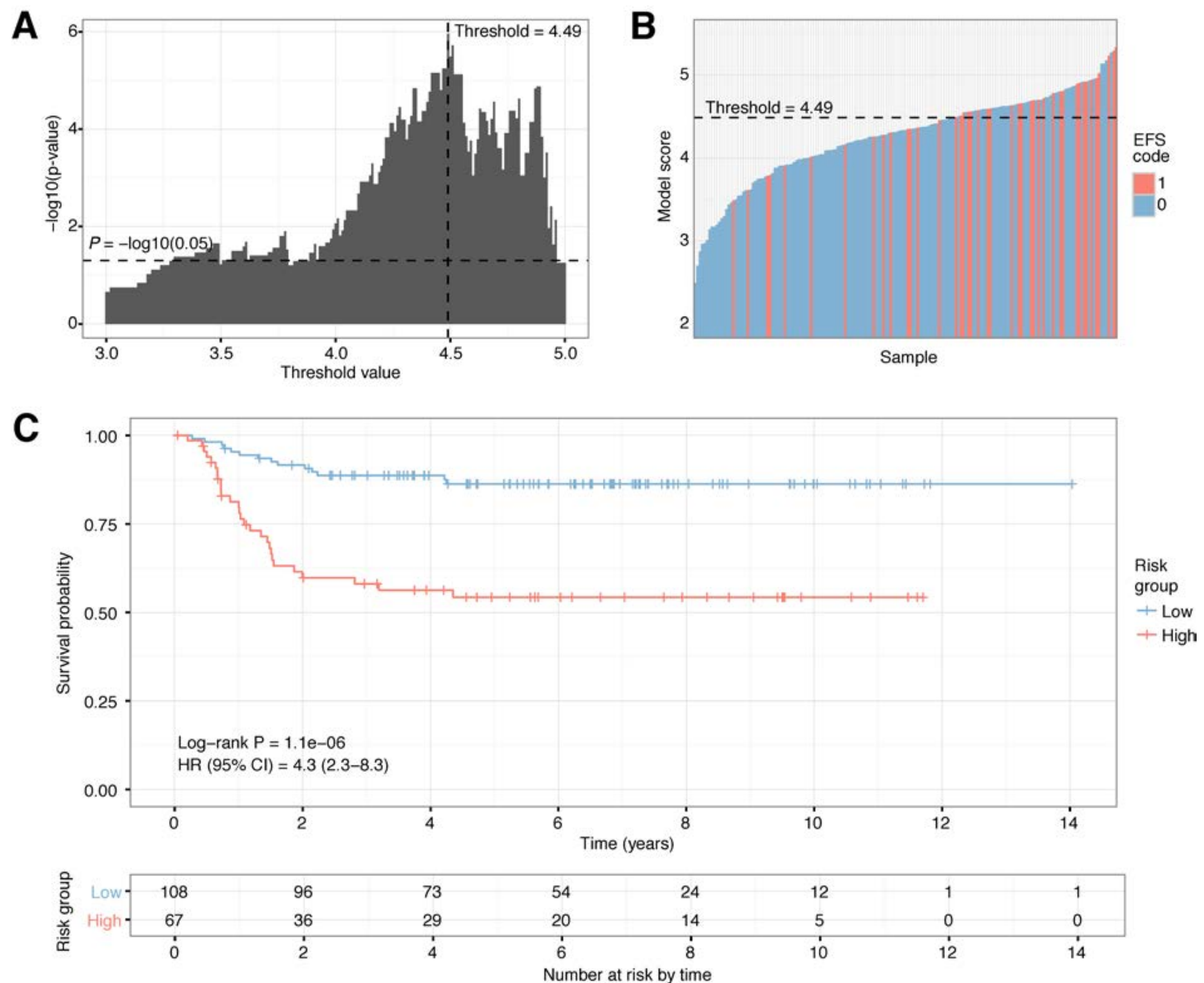
Supplemental Figure 4: Spearman correlation of patient age and cellular component score limited to patients with histological subtype nodular sclerosis across pediatric and adult HL cohorts. Cellular components ordered by Spearman correlation *P*. Red points: pediatric cohort; blue points: adult cohort.



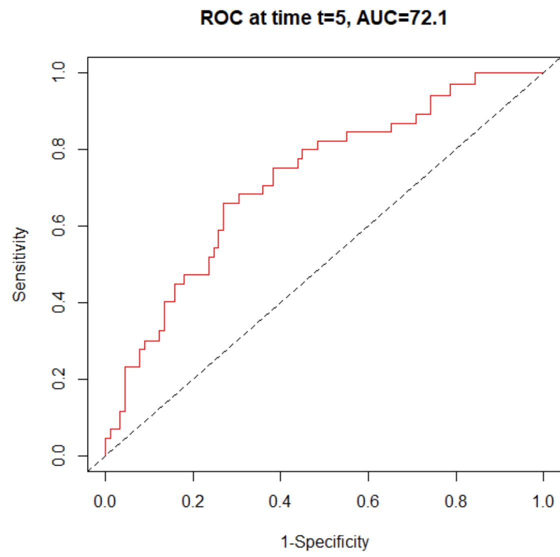
Supplemental Figure 5: Thymus and activation-regulated chemokine (TARC) protein expression and association with survival in pediatrtic HL. (A) Representative examples of two patients from the training cohort showing TARC IHC. Left, TARC low; right, TARC high. **(B)** Kaplan Meier estimates of EFS for patients in the training cohort according to TARC expression as measured by IHC.



Supplemental Figure 6: Genes significantly associated with EFS from univariate Cox regression analysis of the training cohort. The expression levels of 79 genes were significantly associated with EFS ($P < .05$). Three genes achieved an FDR < .1 (indicated).



Supplemental Figure 7: Application of the PHL-9C model to the pediatric HL training cohort. (A) Log-rank p-values of PHL-9C high-risk vs. low-risk using various threshold splits. The x-axis represents the threshold value used to split the training cohort into high-risk and low-risk patients, starting at the minimum PHL-9C model score achieved then incrementing by 0.01 until the maximum PHL-9C model score achieved. The y-axis represents the log-rank p-value for each split. The threshold that obtained the most significant log-rank p-value was chosen (4.49). **(B)** Model scores per patient colour-coded by event status, where EFS code 1 represents event (red), and EFS code 0 represents non-event (blue). Dashed line represents the PHL-9C model score threshold, where patients with model scores below the threshold (4.49) are classified as "low-risk", and patients with model scores above the threshold are classified as "high-risk". **(C)** Kaplan Meier estimates of EFS in the training cohort based on the PHL-9C model score threshold.



Supplemental Figure 8: Time dependent Receiver operating characteristic (ROC) curve for PHL-9C applied to the validation cohort.

REFERENCES

1. Friedman DL, Chen L, Wolden S, et al. Dose-intensive response-based chemotherapy and radiation therapy for children and adolescents with newly diagnosed intermediate-risk hodgkin lymphoma: a report from the Children's Oncology Group Study AHOD0031. *J Clin Oncol*. Nov 2014;32(32):3651-8. doi:10.1200/JCO.2013.52.5410
2. Chan FC, Mottok A, Gerrie AS, et al. Prognostic Model to Predict Post-Autologous Stem-Cell Transplantation Outcomes in Classical Hodgkin Lymphoma. *J Clin Oncol*. Nov 2017;35(32):3722-3733. doi:10.1200/JCO.2017.72.7925
3. Scott DW, Chan FC, Hong F, et al. Gene expression-based model using formalin-fixed paraffin-embedded biopsies predicts overall survival in advanced-stage classical Hodgkin lymphoma. *J Clin Oncol*. Feb 2013;31(6):692-700. doi:10.1200/JCO.2012.43.4589
4. Gray RJ. Weighted analyses for cohort sampling designs. *Lifetime Data Anal*. Mar 2009;15(1):24-40. doi:10.1007/s10985-008-9095-z
5. Viganò E, Gunawardana J, Mottok A, et al. Somatic IL4R mutations in primary mediastinal large B-cell lymphoma lead to constitutive JAK-STAT signaling activation. *Blood*. 05 2018;131(18):2036-2046. doi:10.1182/blood-2017-09-808907
6. Ansell SM, Lesokhin AM, Borrello I, et al. PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N Engl J Med*. Jan 2015;372(4):311-9. doi:10.1056/NEJMoa1411087