Supplementary Methods

Computing 95% confidence regions for (TPR, FPR) and (PPV, NPV).

The details of getting point estimations and 95% confidence regions for paired true positive ratio (TPR), false positive ratio (FPR) and positive predictive value (PPV), negative predictive value (NPV) are described as follows. Patients with success treatment outcomes are assigned as the positive class and patients with failure treatment outcomes are assigned as negative class. A true positive means identifying a treatment success correctly. By definition of (TPR, FPR), the point estimates can be obtained by the following formula.

$\hat{TPR}$ = # of True Positive / N_P

$\hat{FPR}$ = # of False Positive / N_N,

where N_P (N_N) denote the number of positives (negatives). Since the data from the patient with treatment success and treatment failure outcome are independent, we have cov(TPR, FPR)=0. Variance of FPR and FPR can be obtained by the normal approximation to a binomial distribution.

Var($\hat{TPR}$)= $\hat{TPR}(1-\hat{TPR})$/total # of patients,

Var($\hat{FPR}$)= $\hat{FPR}(1-\hat{FPR})$/total # of patients.

Hence, $\begin{pmatrix} TPR \\ FPR \end{pmatrix}$ is asymptotically distributed as Normal $\left(\begin{pmatrix} \hat{TPR} \\ \hat{FPR} \end{pmatrix}, \begin{pmatrix} \text{var}(\hat{TPR}) & 0 \\ 0 & \text{var}(\hat{FPR}) \end{pmatrix}\right)$.

And the 95% confidence region for (TPR, FPR) can be plotted using ellipse package in R.

Similarly, by the definition of (PPV, NPV), their point estimates can be obtained by the following formula.

$\hat{PPV} = \rho\hat{TPR}/\{\rho\hat{TPR} + (1-\rho)\hat{FPR}\}$

$\hat{NPV} = (1-\rho)(1-\hat{FPR})/\{(1-\rho)\hat{FPR} + \rho(1-\hat{TPR})\}$,

where ρ is the treatment success rate of the treatment in the disease population. We used ρ= 0.8 according to (1).

Since we have asymptotic distribution for (TPR, FPR), applying multivariate delta method gives us asymptotic distribution of (logit(PPV), logit(NPV)) as:

Normal $\left(\begin{pmatrix} \log it(\hat{PPV}) \\ \log it(\hat{NPV}) \end{pmatrix}, \begin{pmatrix} \frac{1-\hat{TPR}}{N_p\hat{TPR}} + \frac{1-\hat{FPR}}{N_N\hat{FPR}} & -\frac{1}{N_p} - \frac{1}{N_N} \\ -\frac{1}{N_p} - \frac{1}{N_N} & \frac{\hat{TPR}}{N_p(1-\hat{TPR})} + \frac{\hat{FPR}}{N_N(1-\hat{FPR})} \end{pmatrix}\right)$,

where logit(p)=log(p/(1-p)) for 0<p<1. The detailed proof can be found in (2).

The 95% confidence region for (logit(PPV), logit(NPV)) can be plotted using ellipse package in R. And the 95% confidence interval for (PPV, NPV) can be obtained by take inverse logit transformation of the 95% confidence interval of ((logit(PPV), logit(NPV)).

Reference:

Steidl C, Lee T, Shah SP, Farinha P et al. Tumor-associated macrophages and survival in classic Hodgkin's lymphoma. *N Engl J Med* 2010 Mar 11;362(10):875-85.

Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. 1st ed. Oxford New York; 2003.

Supplementary Table 3: Summary of GI50 Values of NCI 60 cell lines for commonly used drug regimens in cancer treatment

Drug Name	NSC No.	Δ_{GI50} Value	Ten distinct GI50 Value	Applicable to MPI model
5-fluorouracil	19893	2.666	Yes	Yes
adriamycin/doxorubicin	123127	3.338	Yes	Yes
asparaginase	109229	3.553	Yes	Yes
bleomycin	125066	3.361	Yes	Yes
busulfan	750	0.455	Yes	No
carboplatin	241240	0.935	Yes	No
cisplatin	119875	1.616	Yes	Yes
cyclophosphamide	26271	0.451	Yes	No
cytarabine	63878	3.96	Yes	Yes
dacarbazine	45388	1.004	Yes	Yes
daunorubicin	82151	2.89	Yes	Yes
dexamethasone	34521	4	No	No
docetaxel	628503	3.635	Yes	Yes
epirubicin	256942	2.589	Yes	Yes
etoposide	141540	2.707	Yes	Yes
etoposide	141540	2.707	Yes	Yes
goserelin	606864	NA	NA	NA
Idarubicin	256439	2.863	Yes	Yes
megesterol	71423	1.863	Yes	Yes
mercaptopurine	755	3.026	Yes	Yes
methotrexate	740	3.731	Yes	Yes
paclitaxel	125973	3.045	Yes	Yes
prednisolone	63549	3.883	Yes	Yes
rituximab	687451	NA	NA	NA
tamoxifen	180973	0.629	Yes	No
thalidomide	66847	0	No	No
thioguanine	752	2.658	Yes	Yes
valspodar	648265	1.465	Yes	Yes
velcade	681239	1.883	Yes	Yes
vinblastine	49842	3.07	Yes	Yes
vincristine	67574	3.243	Yes	Yes

NSC No.: National Service Center Number

Supplementary Table 4: Point estimates for pairs of (TPR, FPR) and (PPV, NPV) for sensitivity scores with 86th percentiles and mean of all scores cutoff points for study GSE16446

Cutoff point	86th percentile	Mean
(TPR, FPR)	(0.25, 0.12)	(0.63, 0.46)
(PPV, NPV)	(0.25, 0.87)	(0.18, 0.90)

Dichotomized Predictors for Hodgkin's Lymphoma (GSE17920)

We used three different cutoffs to dichotomize SS: 1) the 20th percentile of all scores, which is the historical rate of treatment success for Hodgkin's lymphoma; 2) the 30th percentile of all scores, which is the rate of treatment success for the patients in GSE17920; and 3) the mean of all scores, which is the default cutoff point used by MPI. We summarized the point estimates for (TPR, FPR) and (PPV, NPV) in **Supplementary Tables 5.1 and 5.2**; we plotted their 95% confidence regions in **Supplementary Figure 1A,B**. For comparison purposes, the 95% confidence regions for (TPR, FPR) and (PPV, NPV) for a prediction model using the international prognostic score (IPS) were also plotted in **Supplementary Figure 1**. We used 3 as the cutoff for IPS because historically about 20% percent of Hodgkin's patients have IPS > 3. The dichotomized prediction models using SS alone did not perform better than chance (i.e., the 95% confidence regions overlap the gray region representing the null hypothesis). Using the 30th percentile as the cutoff, 4 patients (out of 130) were reclassified, but the NRI is not greater than zero (**Table 6**; NRI: -0.03, 95% CI: -0.10–0.02).

We used the same cutoffs to dichotomize CS. The point estimates for (TPR, FPR) and (PPV, NPV) pairs are summarized in **Supplementary Tables 5.1 and 5.2**; the 95% confidence regions are plotted in **Supplementary Figure 1C, D**. The 95% confidence regions show that CS performed significantly better than chance (solid blue ellipses and all orange ellipses fall outside the grey areas in panels C and D of **Supplementary Figure 1**). The 95% confidence regions based on a prediction model using IPS>3 are also plotted for comparison. Compared to the predictions based on IPS > 3, however, CS did not perform significantly better (none of the ellipses are totally above the green ellipse in **Supplementary Figure 1**). By this criterion, these models are equivalent to the prediction model based on IPS>3, whose 95% confidence region is also not clearly better than chance. Using the 30th percentile as the cutoff, 18 patients (out of 130) were reclassified comparing predictions based on only clinical features to CS, and the NRI is not greater than zero (**Table 8**; NRI: -0.04, 95% CI: -0.19–0.12).

In addition, since the dotted lines and dotted ellipse in **Supplementary Figure 1** are close to (or overlap) the corresponding solid lines and solid ellipse respectively, and the statistical test results are similar for original scores and adjusted scores (**Table 5** and **Table 6**), the CS and SS based on the original and adjusted gene expression values provide similar prediction results. We conclude that CS and SS are robust to the batch effects present in this instance.

Supplementary Table 5.1: Point estimates for pairs of (TPR, FPR) for sensitive/combined scores based on original GE and adjusted GE with 20 percentiles, 30 percentiles and means of scores cutoff points for study GSE17920

	Based on Original GE			Based on Adjusted GE		
Cutoff point	20th percentile	30th percentile	Mean	20th percentile	30th percentile	Mean
Combined Score	(0.85, 0.66)	(0.78, 0.50)	(0.62, 0.32)	(0.86, 0.66)	(0.78, 0.50)	(0.60, 0.34)
Sensitivity Score	(0.86, 0.74)	(0.75, 0.58)	(0.61, 0.42)	(0.86, 0.74)	(0.75, 0.58)	(0.57, 0.45)

Supplementary Table 5.2: Point estimates for pairs of (PPV, NPV) for sensitive/combined scores based on original GE and adjusted GE with 20 percentiles, 30 percentiles and means of scores cutoff points for study GSE17920

	Based on Original GE			Based on Adjusted GE		
Cutoff point	20th percentile	30th percentile	Mean	20th percentile	30th percentile	Mean
Combined Score	(0.83, 0.37)	(0.86, 0.37)	(0.89, 0.31)	(0.84, 0.38)	(0.86, 0.37)	(0.87, 0.29)
Sensitivity Score	(0.82, 0.27)	(0.84, 0.29)	(0.85, 0.27)	(0.82, 0.27)	(0.82, 0.25)	(0.83, 0.24)

Supplementary Table 6: Treatment Outcomes by different sample centers for study GSE17920 (Hodgkin's Lymphoma)

	Overall (N=130)		Vancouver (N=100)		Nebraska (N=30)		Chi2-test
	N	%	N	%	N	%	P Value
Success	92	70.7	82	82.0	10	33.3	<0.0001
Failure	38	29.3	18	18.0	20	66.7	

Supplementary Table 7: Reclassification Table for the prediction performance without and with the sensitivity score for study GSE17920 (Hodgkin's Lymphoma)*

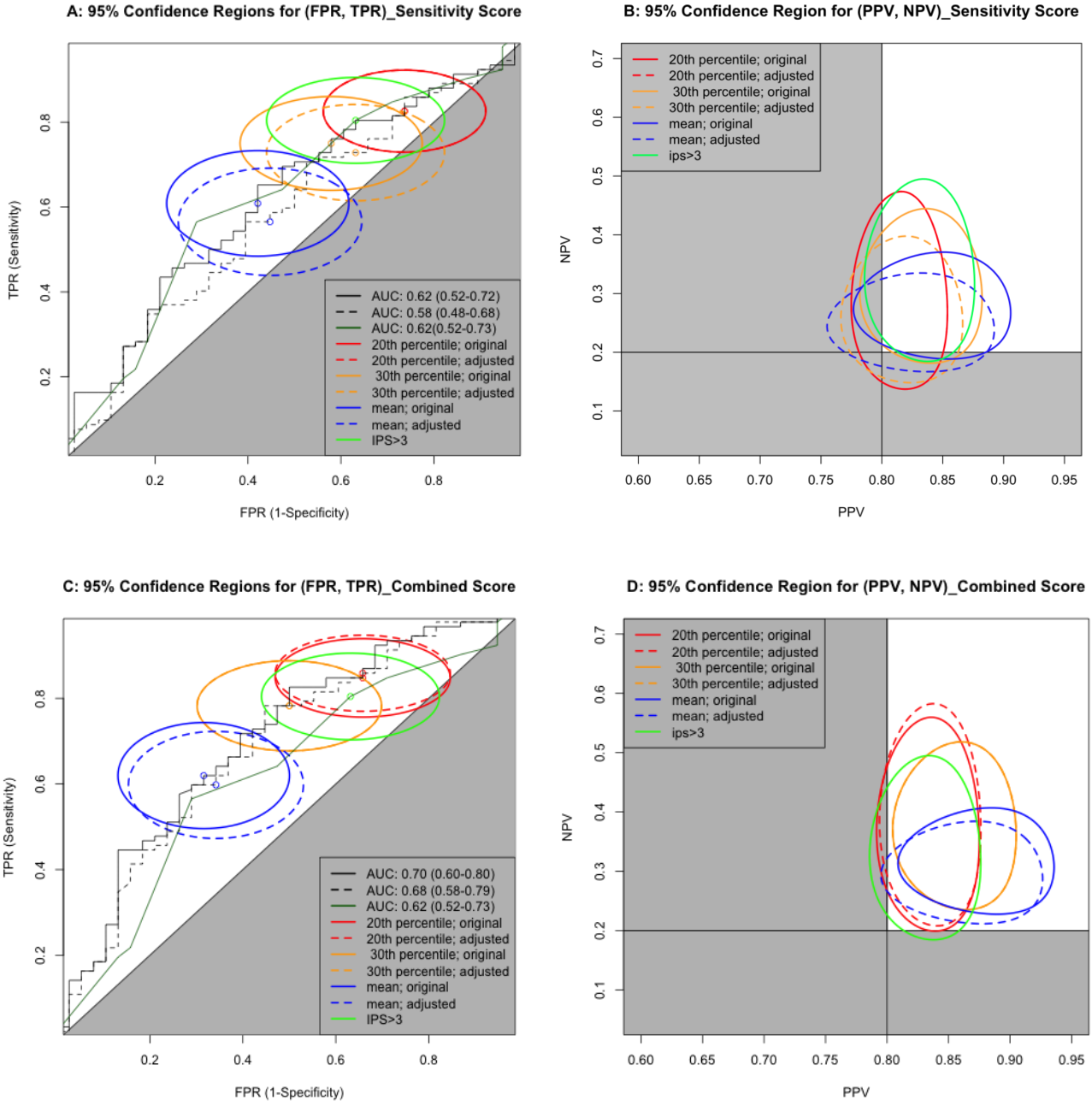
	Age+Stage+Sensitivity Score NRI: -0.03, 95% CI: (-0.10, 0.02)		Combined Score NRI: -0.04, 95% CI: (-0.19, 0.12)		
Age+Stage	**Failure**	**Success**	**Failure**	**Success**	**Total**
Failure	N=37 (N=18 Success)	N=2 (N=1 Success)	N=30 (N=14 Success)	N=9 (N=5 Success)	N=39 (N=19 Success)
Success	N=2 (N=2 Success)	N=89 (N=71 Success)	N=9 (N=6 Success)	N=82 (N=67 Success)	N=91 (N=73 Success)
Total	N=39 (N=20 Success)	N=91 (N=72 Success)	N=39 (N=20 Success)	N=91 (N=72 Success)	N=130 (N=92 Success)

*Cutoff point: 30th percentile; NRI: Net reclassification improvement; CI: confidence interval.

Supplementary Table 8: Univariate Cox regression of MPI sensitivity score related to disease free survival for study GSE10255 (Acute Lymphoblastic Leukemia)*

Parameter	HR	95% CI	P-value
Sensitivity score (continuous)	0.99	(0.97-1.02)	0.75
Sensitivity score (categorized)			
Intermediate vs. Good	0.67	(0.20-2.22)	0.51
Poor vs. Good	1.47	(0.39-5.49)	0.57

*HR: hazard ratio; CI: confidence interval.

Supplementary Figure 1: 95% Confidence Regions of paired (FPR, TPR), Panels A and C, and (PPV, NPV), Panels B and D, for continuous sensitivity and combined score based on original GE (solid ellipses) and adjusted GE (dotted ellipses) with three different cutoff points for study GSE17920 (Hodgkin's Lymphoma)

Cutoff points: 1) 20th percentile (the population response rate, red ellipses); 2) 30th percentile (the response rate of GSE17920 study, orange ellipses) 3) mean (the standard cutoff point used by MPI, blue ellipses).
Green ellipse: the prediction model based on international prognostic score (IPS) greater than 3.
Grey Area: The region for the prediction made by chance.
The solid black line and dotted black line are the ROC curves for prediction score based on original and adjusted gene expressions respectively. The solid green line is the ROC curves for IPS.
FPR: False positive Rate; TPR: True Positive Rate; PPV: Positive Predictive Value; NPV: Negative Predictive Value.
Note: The dotted red ellipse and the solid red ellipse are overlapped for Panel A and B. The dotted orange ellipse and the solid orange ellipse are overlapped for Panel C and D.