Supplementary Online Content

Gao S, Tibiche C, Zou J, et al. Identification and Construction of Combinatory Cancer Hallmark–Based Gene Signature Sets That Predict the Recurrence and Chemotherapy Benefits in of Stage II Colorectal Cancer. *JAMA Oncology*. Published online October 22, 2015. doi:10.1001/jamaoncol.2015.3413.

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods

Data processing and normalization

For Affymetrix data, expression data were processed using the MAS5.0 algorithm. Subsequently, data were log2 transformed, the values of the probes belonging to the same gene were averaged and *z*-score normalized. For Hitachisoft AceGene Human Oligo Chip data, we took the normalized data processed.

For normalizing the metadata derived from Affymetrix arrays, we combined a metadata set for all samples generated using Affymetrix platforms. Affymetrix expression data were processed using the MAS5.0 algorithm. The data were log2 transformed, the values of the probes belonging to the same gene were averaged and then median-centered across arrays. Finally, gene expression data were *z*-score normalized across arrays.

The pseudocode for generating cancer hallmark-based gene signatures using the MSS algorithm.

1. Generate the survival gene pool.
    a. Analyze the gene expression data in the training set for 10-year disease-free survival as implemented previously [1].
    b. Use fuzzy clustering to classify the samples into 2 classes. Genes whose *P*-values are less than a cut off value (P < .05, log-rank test) are regarded as survival genes.
2. Classify the members of the survival gene pool using the functional annotation-clustering tool [2].
    a. Assign survival genes to several gene groups based on selected cancer hallmark related GO (Gene Ontology) terms, such as cell cycle, apoptosis, immunological response and so on, which are closely related to the development and metastasis of cancer. The groups are called GO-term-defined gene sets.
    b. Retain only the gene sets whose size satisfies: size < 100.
3. For each GO-term-defined gene set retained in Step 2, generate 1 million random gene sets (RGSs) each containing 30 genes from the GO-term-defined gene set.
4. Generate *m* (*m*=36 here) random datasets (RDSs) from the training set, maintaining the same ratio of “good” (non-recurred) and “bad” (recurred) tumors as in the original training set.
5. Screen the GO-term-defined gene sets.

```
For each GO-term-defined gene set
    For i=1:m
        For j=1:1000000
            Calculate the P-value of the jth RGS as a signature for survival for the ith RDS
            If (P-value < .05, log-rank test after fuzzy clustering to classify the samples into 2
            classes) p_{i,j} = 1 else p_{i,j} = 0
        End
    End
End
```

6. For each RGS, j, calculate the fraction of RDSs for which it is predictive.
7. Calculate the number of times a gene, k, is a member of an RGS that is predictive at least 90% of the time.

$$Fg_k = \sum_{j}^{1000000} \varphi(k, j)\theta(Fs_j)$$

where $$\varphi(k, j) = \begin{cases} 1 & if\ gene\ k \in RGS\ j \\ 0 & otherwise \end{cases}$$

and $$\theta(Fs_j) = \begin{cases} 1 & if\ Fs_j > 0.9 \\ 0 & otherwise \end{cases}$$

8. For each GO-term-defined gene set, rank the genes in the set according to their Fg_k and retain the top 30.

Comments on MSS

More description of the MSS. For the recurred and non-recurred samples of the training set, we applied fuzzy clustering to classify the samples into 2 classes and then conducted log-rank test to indentify modulated genes (P < .05). From these modulated genes, we collected hallmark GO-defined genes. For a cancer hallmark GO-defined genes, we narrowed down to 60-100 genes. From the training set, we generated 36 random datasets, at the same time, we generated 1 million of random gene sets (each set contains 30 genes from the GO-defined genes). We conducted fuzzy clustering (k=2) for each random dataset using each random gene set (1 million of random gene sets in total) to distinguish low- and high-risk groups (log-rank test, P < .05). For a given cancer hallmark GO term, if 1,000 – 5,000 of random gene sets (called passed gene-sets) are able to distinguish low- and high-risk groups (log-rank test, P <.05) in more than 90% of the 36 random datasets (P < .005), we collected and ranked the genes based on their frequency among the passed gene-sets. The top ranked genes (30 genes) can be used as a gene signature for that cancer hallmark. The cutoff of 30 is based on a systemic analysis of MSS-derived ranked genes (see details in the Supplementary Materials of Li et al., 2010[3]).

How to determine which hallmark-based GO term could generate a gene signature. For a given cancer hallmark-based GO term, the number of the passed gene-sets is greater than 1,000 but less than 5,000 (they are from 1 million of random gene sets, P <.005). If it is less than 1,000, the ranked genes could not be robust. If the number of the passed gene-sets is greater than 5,000, one could extract a subsets of the passed gene-sets by lowing down P value cutoff.

Robustness of the MSS-derived gene signatures. Three components of the MSS make the gene signatures more robust than other methods: (1) higher genome instability in cancer cells makes tumor cells often have many 'passenger signals' than other types of cells, which means that the variability of gene expression profiles between individual tumors can be extremely high, and the 'real' cancer gene expression signals may be buried in these highly varied profiles. To cope with this problem, in MSS we focused on cancer hallmark associated genes which are more likely associated with cancer progression and metastasis and less likely 'passenger signals'. Therefore, the gene signatures derived from cancer hallmark associated genes are most likely real signals and will be more robust in independent data sets; (2) In MSS, we asked that ‘passed gene-sets’ which are able to distinguish recurred and non-recurred samples among 90% of the 36 random datasets which were generated from the original training set. Therefore, the selected ‘passed

gene-sets' have higher robustness; (3) Finally, we collected the 1,000 - 5,000 passed gene-sets from 1 million of random gene sets (P <.005), and then ranked these sets (eg, assuming that high frequent genes in the passed gene-sets may contribute more for distinguishing recurred and non-recurred samples). By doing so, the gene signatures are more robust.

Compare the MSS-derived signatures with other signatures in terms of robustness. First of all, the stage II CRC gene signatures in this study are predictive in 11 independent cohorts. In a robustness test of colon cancer prognostic gene signatures conducted by Park et al., three out of five tested gene signatures failed to predict prognosis in two independent validation sets[4]. They are (1) Veridex 7-gene relapse hazard score (V7RHS) developed by conducting univariate Cox proportional hazards regression followed by *t* test to construct relapse hazard scores[5]; (2) Meta163 (metastasis-associated 163-gene expression signature) developed by unsupervised hierarchical clustering[6]; (3) 13-gene ColoGuideEx prognostic predictor developed by (lasso) penalized Cox proportional hazards analyses[7]. Of note, the study conducted by Park et al. had tested only two independent validation sets to evaluate the robustness of the gene signatures. In our study, we tested the CSS sets in 11 independent cohorts and showed that they worked well.

Leave-one-out cross-validations using cancer-hallmark-based gene signature sets (CSS sets)

For the samples which have not been treated with drugs in the 11 independent cohorts, leave-one-out cross-validations were applied. Each sample was cross-validated by each gene signature of a CSS set using the predicting algorithm described previously [3]. The pseudocode for leave-one-out cross-validations using CSS sets:

For *i*=1..n (n is the total number of samples of stage I and II)

1. Dividing samples in the dataset into two groups. The *i*th sample is in Group 1, other n-1 samples are in Group 2.
2. Extracting feature vectors from gene signatures. For the *i*th sample, its vector is V_{ij}= (g_{1ij}, g_{2ij}, …, g_{kij}). g_{kij} is the expression value of the *k*th gene of the *j*th gene signature (GS-j) in the *i*th sample (k=1, 2, … 30, j=1, 2, 3,…8). Feature vector V_{ijg} is shrunken class centroids extracted from the relapse free patient samples from Group 2 for GS-j using PAMR method [8] and V_{ijb} is shrunken class centroids extracted from the metastasis patient samples from Group2 for GS-j.
3. Classifying the *i*th sample using the feature vectors. Assigning the *i*th sample to the relapse free group if Cor(V_{ijg},V_{ij})>=Cor(V_{ijb},V_{ij}), otherwise to the metastasis group (Cor(V_{ijg},V_{ij}) is the Pearson correlation coefficient between V_{ijg} and V_{ij}). Only the predictions of stage II samples were used for the following analyses.
4. Classifying the *i*th sample by combining the outcomes from step 3. Assigning the *i*th sample to the low-risk group if it is classified in the relapse free group by four or more gene signatures, or to the high-risk group if it is assigned to the metastasis group by all eight signatures, otherwise to the intermediate-risk group.

End

Constructing combinatory cancer-hallmark-based gene signature sets (CSS sets) for predicting prognosis

To determine the number of gene signatures for a CSS set, we systematically tested combinatory predictions using N (N=1, 2, 3…) of MSS-derived gene signatures in the pooled set of GSE37892 and GSE33113 (eTable 20). N is defined as the number of gene signatures in a CSS set to reach a higher prediction accuracy and a relatively higher recall rate for stage II samples (see eTable 19 and eTables 22-24). When N is set, the predictive accuracy could not be significantly improved or

recall rate could be significantly dropped by adding one or more additional gene signatures to the CSS set.

In general, collaborative cancer hallmark gene signatures within a CSS set could foster greater cohesion and interaction and therefore could increase prediction accuracy and boost the recall rate by unifying the predictions derived from multiple CSS sets. The recall rate was defined as the fraction of recurrent (or non-recurrent) samples that have been called in the predicted high-risk (or low-risk) patient groups. For different samples of a cancer type or subtype, distinct gene sets could be used for the molecular mechanism of tumor recurrence. Therefore, each CSS set predicts a fraction of the recurrent samples. By merging the predictions derived from multiple CSS sets, the overall recall rate is boosted. In summary, the CSS set approach uses an intersection function (i.e., consensus prediction) to improve the accuracy of identifying individual samples and then uses a union function to improve the overall recall rate for prediction.

The pseudocode for determining the number of gene signatures for a CSS set:

For *I*=1..m (m is the total number of gene signatures, here is eight)
  For *i*=1:n (n is the total number of samples of stage I and II)
    For *j*=1:m
      Assign the *i*th sample to relapse free or metastasis group based on the *j*th gene signature (see details in the pseudocode of section Leave-one-out cross-validations using CSS sets)
    End
    If (the number of assignments for relapse free>=*I* & the *i*th sample is stage II) the *i*th sample is replase free and assigned to Group RF
    If (the number of assignments for metastasis>=*I* & the *i*th sample is stage II) the *i*th sample is metastasis and assigned to Group M
  End
  Calculate the $Accuracy_I$ and $Recall_I$ for Group RF and Group M respectively
End
Select N as *I* reaching balance between prediction accuracy and recall rate (i.e., looking for high accuracy and relatively higher recall rate, see eTable 19 and eTables 22-24) for Group RF and Group, M respectively

Assigning samples with gene signature scores (GSSs)

For each sample in the validation sets, we counted the number of gene signatures which predicted the sample to be low-risk. For example, if a sample has none of the eight gene signatures predicted to be low-risk, we assigned a score 0 to that sample. On the other hand, if a sample has eight gene signatures predicted to be low-risk, we assigned a score 8 to that sample.

$$GSS_i = \sum_{j=1}^{n} f_{low-risk}(GS-j,i), n = 8$$

$$f_{low-risk}(GS-j,i) = \begin{cases} 1 & if\ Cor(V_{ijg},V_{ij}) >= Cor(V_{ijb},V_{ij}) \\ 0 & otherwise \end{cases}$$

GS-j is the *j*th gene signature, V_{ij} is the expression value of GS-j in the *i*th sample, V_{ijg} and V_{ijb} are shrunken class centroids extracted from the relapse free patient samples and metastasis samples respectively (the *i*th sample is not included) using PAMR method [8]. $Cor(V_{ijg},V_{ij})$ is the Pearson correlation coefficient between V_{ijg} and V_{ij}.

Examine the robustness of CSS sets when pooling the training set with other different datasets

In the main text, to determine the number of gene signatures for a CSS set, we used the stage II samples pooled from GSE37892 (the training set) and GSE33133. To examine the potential effects of this process by using different datasets, we applied the same CSS building process using the stage II samples pooled from GSE37892 (the training set) and other cohorts. To be pooled with the data of GSE37892 (the training set), the tested datasets should use the same microarray platform of GSE37892, and furthermore, contain follow-up clinical information of the patients. Based on these criteria, GSE17538, GSE21510 and GSE27854, respectively, were used to pool with GSE37892 to determine the number of gene signatures for CSS sets (eTables 22-24). To select the optimal number of the gene signatures for constructing CSS sets, we balanced the predicting accuracy and recall rate (accuracy should be higher, recall rate is relatively higher too), therefore, for predicting low-risk patients, any four of the eight signatures is good enough (eTables 22-24). This conclusion is in agreement with that determined in the main text (pooling of GSE37892 and GSE33133, see eTable 19). For predicting high-risk patients, either all the eight signatures (eTable 22), which is the same as that determined in the main text, eTable 19), or any seven of the eight signatures are good enough (eTables 23-24).

It is noted that the results of eTable 19 are more robust than the results from others (eTables 22-24), because the recurred samples in GSE33133 (eTable 19) are much more than those in other 3 datasets (eTables 22-24) where the recurred sample sizes are small. Nevertheless, we tested the predictions using the CSS sets (any four of the eight signature for predicting low-risk patients, and any seven of the eight signatures for predicting high-risk patients) derived from eTables 23-24. As shown in eTable 25 and eFigures 3-4, similar results were obtained when pooling different datasets with the training set to construct CSS sets. The predictions of prognosis and the survival benefit of 5-FU based therapy were still true when using these CSS sets derived from the pooling of GSE37892 (the training set) with GSE33133, GSE17538, GSE21510 or GSE27854, respectively. These results confirmed the robustness of the CSS set approach.

eTable 1 Patient clinical characteristics for the training and validation sets

Variable	Training set (n=162)		Validation set (n=843)	
Clinical Characteristic	Number of patients	%	Number of patients	%
Age, years				
Median	71.9		69	
≤70	75	46.3	330	39.1
＞70	86	53.1	253	30.0
NA	1	0.6	260	30.8
Gender				
Female	81	50.0	258	30.6
Male	81	50.0	302	35.8
NA	0	50.0	283	33.6
Localization				
Left	40	24.7	103	12.2
Right	32	19.8	68	8.1
Rectum	0	0.0	23	2.7
NA	90	55.6	649	77.0
Relapse				
No	136	84.0	665	78.9
Yes	26	16.0	178	21.1
Median time to relapse, months	14.8		22.0	
Median follow-up, months	63.5		53.0	

eTable 2 Patient follow-up time information of cohorts

GEO	GSE37892	GSE33113	GSE26096	GSE17538	GSE39582	GSE21510	GSE14333
Follow-up time	Yes	Yes	No	Yes	Yes	No	Yes
GEO	GSE27854	GSE12945	GSE41258	GSE16125	GSE24551	GSE12032	
Follow-up time	No	Yes	Yes	No	Yes	No	

Note: all the cohorts contain relapse or metastasis information. Among them, some contain follow-up time information, but some do not.

eTable 3 Patient clinical characteristics for the training and validation sets

Variable Clinical Characteristic	Training set (n=162) Number of patients	%	Validation set (n=843) Number of patients	%
Chemotherapy				
No			340	40.3
Yes	NA		76	9.0
NA			427	50.7
Relapse location				
Distal			155	18.4
Proximal	NA		97	11.5
NA			591	70.1
Metastasis				
No			68	8.1
Yes	NA		22	2.6
NA			753	89.3
Grade				
1			5	0.6
2	NA		65	7.7
3			16	1.9
NA			757	89.8
Death				
No			239	28.4
Yes	NA		83	9.8
NA			521	61.8
Median time to death, months			37	

Variable Clinical Characteristic	Training set (n=162) Number of patients	%	Validation set (n=843) Number of patients	%
pT				
2			4	0.5
3	NA		258	30.6
4			52	6.2
NA			529	62.8
pN				
0	NA		314	37.2

1		0	0.0
NA		529	62.8
pM			
0		314	37.2
1	NA	0	0.0
NA		529	62.8
APC mutation			
No	NA	16	1.9
Yes		79	9.4
NA		748	88.7
BRAF mutation			
No		215	25.5
Yes	NA	20	2.4
NA		608	72.1
KRAS mutation			
No		166	19.7
Yes	NA	83	9.8
NA		594	70.5
TP53 mutation			
No		91	10.8
Yes	NA	90	10.7
NA		662	78.5
MMR Status			
dMMR		35	4.2
pMMR	NA	177	21.0
NA		631	74.9
CIMP Status			
−		185	21.9
+	NA	43	5.1
NA		615	73.0
CIN Status			
−	NA	57	6.8
+		156	18.5

Variable Clinical Characteristic	Training set (n=162)		Validation set (n=843)	
	Number of patients	%	Number of patients	%

NA		630	74.7
MSI Status			
MSS	NA	89	10.6
MSI-low		9	1.1
MSI-high		22	2.6
NA		723	85.8

eTable 4 Patient clinical characteristics for the training set, GSE37892

Clinical Characteristic	Number of patients	%
Age, years		
Median	71	
≤70	36	49.3
＞70	37	50.7
Gender		
Female	34	46.6
Male	39	53.4
Localization		
Left	40	54.8
Right	32	43.8
NA	1	1.4
Relapse		
No	65	89.0
Yes	8	11.0
Median time to relapse, months	21.2	
Median follow-up, months	51.1	

eTable 5 Patient clinical characteristics for the validation set, GSE12032

Clinical Characteristic	Number of patients	%
Relapse		
No	62	67.4
Yes	30	32.6

eTable 6 Patient clinical characteristics for the training set, GSE33113

Clinical Characteristic	Number of patients	%
Age, years		
Median	73.7	
≤70	39	43.8
＞70	49	55.1
NA	1	1.1
Gender		
Female	47	52.8
Male	42	47.2
Relapse		
No	71	79.8
Yes	18	20.2
Median time to relapse, months	14.6	
Median follow-up, months	73.7	

eTable 7 Patient clinical characteristics for the validation set, GSE27854

Clinical Characteristic	Number of patients	%
Relapse		
No	34	82.9
Yes	7	17.1

eTable 8 Patient clinical characteristics for the validation set, GSE21510

Clinical Characteristic	Number of patients	%
Relapse		
No	31	83.8
Yes	6	16.2

eTable 9 Patient clinical characteristics for the validation set, GSE26906

Clinical Characteristic	Number of patients	%
Age, years		
Median	69.5	
≤70	53	58.9
＞70	37	41.1
Gender		
Female	47	52.2
Male	43	47.8
Localization		
Left	66	73.3
Right	24	26.7
Metastasis		
No	68	75.6
Yes	22	24.4
Apc mutation		
No	14	15.6
Yes	76	84.4
Relapse		
No	69	76.7
Yes	21	23.3

eTable 10 Patient clinical characteristics for the validation set, GSE17538

Clinical Characteristic	Number of patients	%
Age, years		
Median	69.5	
≤70	38	54.3
＞70	32	45.7
Gender		
Female	36	51.4
Male	34	48.6
Grade		
1	5	7.1
2	52	74.3
3	6	8.6
NA	7	10.0
Death		
No	58	63.0
Yes	34	37.0
Median time to death, months	32.1	
Relapse		
No	59	85.3
Yes	11	15.7
Median time to relapse, months	22.3	
Median follow-up, months	48.7	
Chemotherapy		
No	62	88.6
Yes	8	11.4

eTable 11 Patient clinical characteristics for the validation set, GSE39582

Clinical Characteristic	Number of patients	%
Age, years		
Median	69	
≤70	140	55.6
＞70	112	44.4
Gender		
Female	102	40.5
Male	150	59.5
Relapse location		
Distal	155	61.5
Proximal	97	38.5
Death		
No	181	82.3
Yes	39	17.7
Median time to death, months	39	
Relapse		
No	196	77.8
Yes	56	22.2
Median time to relapse, months	17	
Median follow-up, months	59.5	
Chemotherapy		
No	198	78.6
Yes	54	21.4
pT		
2	4	1.6
3	189	75.0
4	49	19.4
NA	10	4.0
pN		
0	242	96.0
1	0	0.0
NA	10	4.0

Clinical Characteristic	Number of patients	%
pM		
0	242	96.0
1	0	0.0
NA	10	4.0
KRAS mutation		
No	163	64.7
Yes	81	32.1
NA	8	3.2
BRAF mutation		
No	215	85.3
Yes	20	7.9
NA	17	6.7
TP53 mutation		
No	76	30.2
Yes	68	27.0
NA	108	42.9
MMR Status		
dMMR	35	13.9
pMMR	177	70.2
NA	40	15.9
CIMP Status		
−	185	73.4
+	43	17.1
NA	24	9.5
CIN Status		
−	57	22.6
+	156	61.9
NA	39	15.5

Abbreviations: MMR, mismatch repair; dMMR, deficient MMR; pMMR, proficient MMR; CIMP; CIN, chromosomal instability.

eTable 12 Patient clinical characteristics for the validation set, GSE14333

Clinical Characteristic	Number of patients	%
Age, years		
Median	70	
≤70	80	85.1
＞70	14	14.9
Gender		
Female	45	47.9
Male	49	52.1
Localization		
Left	37	39.8
Right	44	47.3
Rectum	12	12.9
Relapse		
No	80	85.1
Yes	14	14.9
Median time to relapse, months	20.9	
Median follow-up, months	37.7	
Chemotherapy		
No	72	76.6
Yes	22	23.4

eTable 13 Patient clinical characteristics for the validation set, GSE12945

Clinical Characteristic	Number of patients	%
Age, years		
Median	66	
≤70	16	69.6
＞70	7	30.4
Localization		
Rectum	11	47.8
Colon	12	52.2
Grade		
1	0	0.0
2	13	56.5
3	10	43.5
Relapse		
No	22	95.7
Yes	1	4.3
Median time to relapse, months	39	
Median follow-up, months	53.0	
pT		
2	0	0.0
3	23	100.0
4	0	0.0
pN		
0	23	100.0
1	0	0.0
pM		
0	23	100.0
1	0	0.0

eTable 14 Patient clinical characteristics for the validation set, GSE41258

Clinical Characteristic	Number of patients	%
Age, years		
Median	68	
≤70	31	63.3
＞70	18	36.7
Gender		
Female	25	51.0
Male	24	49.0
Relapse		
No	45	91.8
Yes	4	8.2
Median time to relapse, months	117.5	
Median follow-up, months	86.0	
pT		
2	0	0.0
3	46	93.9
4	3	6.1
pN		
0	49	100.0
1	0	0.0
pM		
0	49	100.0
1	0	0.0
TP53 mutation		
No	14	28.6
Yes	18	36.7
NA	17	34.7
MSI Status		
MSS	25	51.0
MSI-low	5	10.2
MSI-high	10	20.4
NA	9	18.4

eTable 15 Patient clinical characteristics for the validation set, GSE16125

Clinical Characteristic	Number of patients	%
Age, years		
Median	65	
≤70	3	60.0
＞70	2	40.0
Gender		
Female	3	60.0
Male	2	40.0
Relapse		
No	10	100.0
Yes	0	0.0
KRAS mutation		
No	3	60.0
Yes	2	40.0
TP53 mutation		
No	1	20.0
Yes	4	80.0
APC mutation		
No	2	40.0
Yes	3	60.0

eTable 16 Patient clinical characteristics for the validation set, GSE24551

Clinical Characteristic	Number of patients	%
Relapse		
No	62	68.9
Yes	28	31.1
Median time to relapse, months	22.7	
Median follow-up, months	47.8	
MSI Status		
MSS	64	71.1
MSI-low	4	4.4
MSI-high	12	13.3
NA	10	11.1

eTable 17 Selected cancer hallmark-associated Gene Ontology terms

GO term	Cancer hallmark[9]	Cancer hallmark network[10]
Cell cycle	Sustaining proliferating signaling, Enabling replicating immortality	Cancer survival network
Cell motility	Activating invasion and metastasis	EMT-network, Stroma-network
Immune response	Avoiding immune destruction	Immune-escaping network, Stroma-network
DNA repair	Genome instability and mutation	Mutation network
Apoptosis	Resisting cell death	Cancer survival network
Cell death	Resisting cell death	Cancer survival network
Phosphorylation	Most of the cancer hallmarks	Most of the cancer hallmark networks

eTable 18 List of cancer hallmark-based gene signatures and their genes

Apoptosis	Cell Cycle	Cell Death	Cell Motility	DNA Repair	Immune Response	Phosphorylation 1	Phosphorylation 2
GRIK2	RGS14	GRIK2	SCYL3	MUS81	NFKB2	SCYL3	PDE6H
NLRP3	TEX11	KCNC3	C14orf104	EYA4	LILRA4	FES	LPAR2
EPO	TDRD1	UBE4B	HMGCR	HSPA1L	EBI3	EPHB6	FLT4
UBE4B	CTNNB1	TGM2	SMO	APTX	S1PR4	CAMK4	NPR1
FOXL2	PA2G4P4	FOXL2	SELPLG	UPF1	SPN	RIOK2	ADRA2B
TRAF1	SMARCB1	CIAPIN1	EDN2	TREX1	KIR2DS5	TNIK	SRC
NME6	TUBB1	MCF2	SCNN1B	NEIL1	APOA4	MAPK7	CIT
ITGB2	NPAT	ZMAT3	RPS6KB1	SHFM1	KIR2DS1	MUL1	MUL1
PPP3R1	CUL3	PAX3	CXCL3	RAD17	CRISP3	RPS6KB1	F2
PAX3	ZMYND11	CDKN2D	PEX5	SFPQ	PGLYRP1	ALPK3	GRK6
MCF2	CDKN2D	PPP3R1	ASTN1	ATXN3	BCL3	RPS6KA2	NTRK1
PTK2B	CIT	NME6	VAV3	CRY1	IL2	ERC1	CCL11
CTNNB1	KATNB1	JUN	CORO1A	DCLRE1C	VTCN1	BCKDK	PIK3CB
JUN	EGFR	TP53AIP1	LTB4R2	APEX1	IL9	TWF1	OSM
OSM	CCNE1	TRAF1	TNF	UNG	FCGRT	TYRO3	PTK2B
TNFSF13	TNFSF13	DAXX	TNFSF13	TNFSF13	TNFSF13	TNFSF13	TNFSF13
CIAPIN1	RASSF4	XIAP	NDE1	ANKRD17	PSEN1	CAMK1D	EPHB6
FAF1	INCENP	BAX	B4GALT1	RAD23B	EXOSC9	MAP3K14	PLA2G1B
BAX	DAXX	PSEN1	TGFBR1	SMC5	MAVS	MAPKAPK2	STK38L
SIRT1	ANAPC13	PKM2	SHH	DDB2	IL1RL2	TOP1	CDC42BPB
SLC25A6	DMC1	OSM	NCK1	SOD2	CMKLR1	SMAD7	PSEN1
RPS3	PARD6A	KRT20	PTENP1	XPC	HLA-DOA	CDK16	PNKP
MUL1	IL8	SLK	GFRA3	XAB2	LILRB4	IKBKB	PIK3R1
SLK	BANP	ATM	SMCP	SIRT1	PGLYRP4	LMTK2	DAPK1
COL4A3	PSMC4	MUL1	PSG2	MGMT	CPLX2	TNK1	CDK17
TP53AIP1	MAP3K11	MAP3K11	SCARB1	CCNO	PRG2	AGK	SLK
PSEN1	MIS12	MUC5AC	SYK	RAD52	MLH1	TGFB2	PSKH1
DAXX	PCNP	FAIM	PAFAH1B1	CEP164	OSM	TRIM24	PRKCA
FAIM	RBBP4	MYC	PTK2B	POLD4	LYST	NDUFA2	PTPN11
ELMO2	SKP1	ELMO2	YWHAE	UVRAG	AZGP1P1	CSNK2A2	EGFR

eTable 19 Constructing CSS sets by pooling of GSE37892 and GSE31133

Number of Signatures	Low-risk		High-risk	
	Accuracy	Recall	Accuracy	Recall
1	0.906475	0.926471	0.204918	0.961538
2	0.921875	0.867647	0.237113	0.884615
3	0.940171	0.808824	0.2875	0.884615
4	0.953704	0.757353	0.318841	0.846154
5	0.956989	0.654412	0.388889	0.807692
6	0.963415	0.580882	0.422222	0.730769
7	0.953846	0.455882	0.470588	0.615385
8	0.975	0.286765	0.565217	0.5

eTable 20 Array platform and sample numbers of the datasets

Data set	Array Platform	Number of samples	Non-relapsed sample	Relapsed sample
GSE37892	Affymetrix HG-U133	73	65	8
GSE33113	Affymetrix HG-U133	89	71	18
GSE26906	Affymetrix HG-U133	90	69	21
GSE17538	Affymetrix HG-U133	70	59	11
GSE39582	Affymetrix HG-U133	252	196	56
GSE21510	Affymetrix HG-U133	37	31	6
GSE14333	Affymetrix HG-U133	94	80	14
GSE27854	Affymetrix HG-U133	41	34	7
GSE12945	Affymetrix HG-133A	23	22	1
GSE41258	Affymetrix HG-133A	49	45	4
GSE16125	Affymetrix Human Exon 1.0	10	10	0
GSE24551	Affymetrix Human Exon 1.0	90	62	28
GSE12032	Hitachisoft AceGene Human Oligo Chip	92	62	30

eTable 21 Univariate analysis for relapse-free survival in validation sets

Variable	*P*	HR	95% CI
Age, ≤ *v* > 70 years	0.667	1.1	0.71-1.7
Localization, right *v* left	0.4	0.58	0.16-2.06
Gender, male *v* female	0.58	1.13	0.74-1.73
pT, continuous	0.001	2.63	1.46-4.73
pT, T4 *v* T3	0.002	2.6	1.44-4.71
BRAF mutation, M *v* WT	0.43	1.46	0.58-3.68
TP53 mutation, M *v* WT	0.46	1.25	0.7-2.23
KRAS mutation, M *v* WT	0.28	1.35	0.79-2.3
MMR status, dMMR *v* pMMR	0.54	0.77	0.33-1.8
CIMP status, + *v* -	0.97	1.02	0.47-2.18
CIN status, + *v* -	0.2	1.78	0.74-4.25

Abbreviations: HR, hazard ratio; CI, confidence interval; MMR, mismatch repair; dMMR, deficient MMR; pMMR, proficient MMR; CIMP; CIN, chromosomal instability.

eTable 22 Constructing CSS sets by pooling of GSE37892 and GSE17538 (drug treated samples were removed)

Number of Signatures	Low-risk		High-risk	
	Accuracy	Recall	Accuracy	Recall
1	0.912	0.966102	0.155963	1
2	0.909836	0.940678	0.190476	0.941176
3	0.927273	0.864407	0.234375	0.882353
4	0.957447	0.762712	0.254902	0.764706
5	0.952381	0.677966	0.317073	0.764706
6	0.971831	0.584746	0.36	0.529412
7	0.980392	0.423729	0.461538	0.352941
8	1	0.220339	0.6	0.352941

eTable 23 Constructing CSS sets by pooling of GSE37892 and GSE21510 (drug treated samples were removed)

Number of Signatures	Low-risk		High-risk	
	Accuracy	Recall	Accuracy	Recall
1	0.880734	1	0.154762	0.928571
2	0.895238	0.979167	0.166667	0.857143
3	0.934066	0.885417	0.189655	0.785714
4	0.932432	0.71875	0.22	0.785714
5	0.95	0.59375	0.25	0.642857
6	0.942308	0.510417	0.421053	0.571429
7	0.947368	0.375	0.6	0.214286
8	0.961538	0.260417	1	0.071429

eTable 24 Constructing CSS sets by pooling of GSE37892 and GSE27854 (drug treated samples were removed)

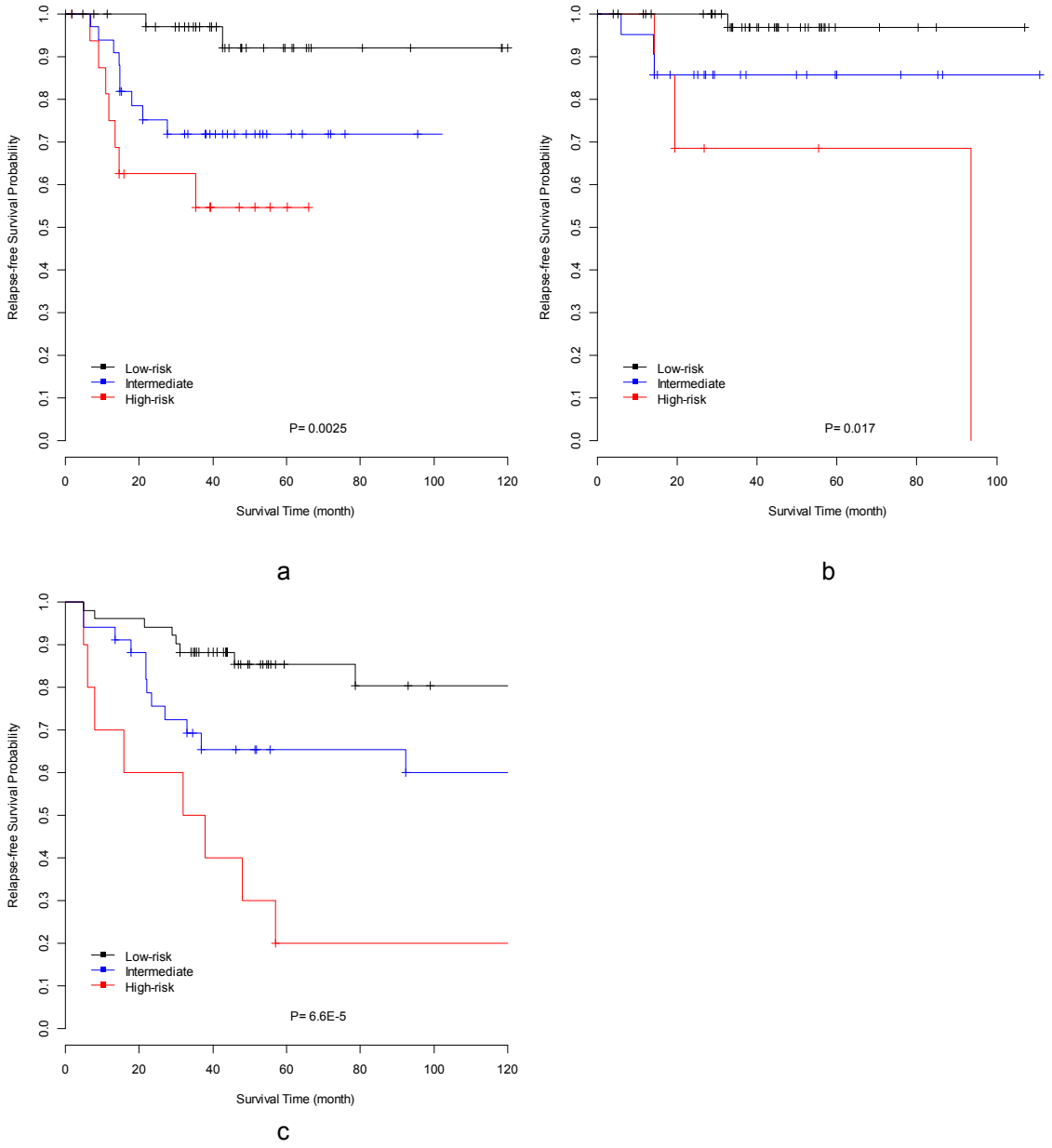
Number of Signatures	Low-risk		High-risk	
	Accuracy	Recall	Accuracy	Recall
1	0.883929	1	0.170732	0.933333
2	0.903846	0.949495	0.188406	0.866667
3	0.923077	0.848485	0.2	0.8
4	0.946667	0.717172	0.230769	0.8
5	0.951613	0.59596	0.282051	0.733333
6	0.944444	0.515152	0.347826	0.533333
7	0.955556	0.434343	0.5	0.333333
8	0.96875	0.313131	1	0.133333

eTable 25 Prediction accuracies and recall rates for stage II CRC patients using the CSS sets constructing rules derived from eTables 23-24

Dataset	Number of samples	Low-risk		Intermediate-risk		High-risk	
		Accuracy*	Recall†	Accuracy*	Recall†	Accuracy**	Recall††
GSE37892 (training)	73	96.2%	78.5%	85.7%	10.8%	38.5%	62.5%
GSE33113	90	94.7%	50.7%	73.7%	19.7%	34.4%	61.1%
GSE26906	90	83.8%	44.9%	66.7%	26.1%	23.1%	28.6%
GSE17538	72	100.0%	66.0%	1.0%	17.0%	50.0%	100.0%
GSE39582	264	89.5%	74.8%	84.6%	13.8%	53.8%	53.8%
GSE21510	54	100.0%	93.5%	100.0%	0.03%	85.7%	100.0%
GSE14333	94	97.7%	64.6%	94.1%	24.6%	41.7%	71.4%
GSE27854	41	100.0%	79.4%	100.0%	17.6%	87.5%	100.0%
GSE12945 and GSE41258	72	93.8%	67.2%	0.0%	0.0%	NA	NA
GSE16125 and GSE24551	95	84.3%	64.2%	72.0%	26.9%	68.4%	46.4%
GSE12032	92	88.1%	95.2%	25.0%	0.05%	100.0%	43.3%

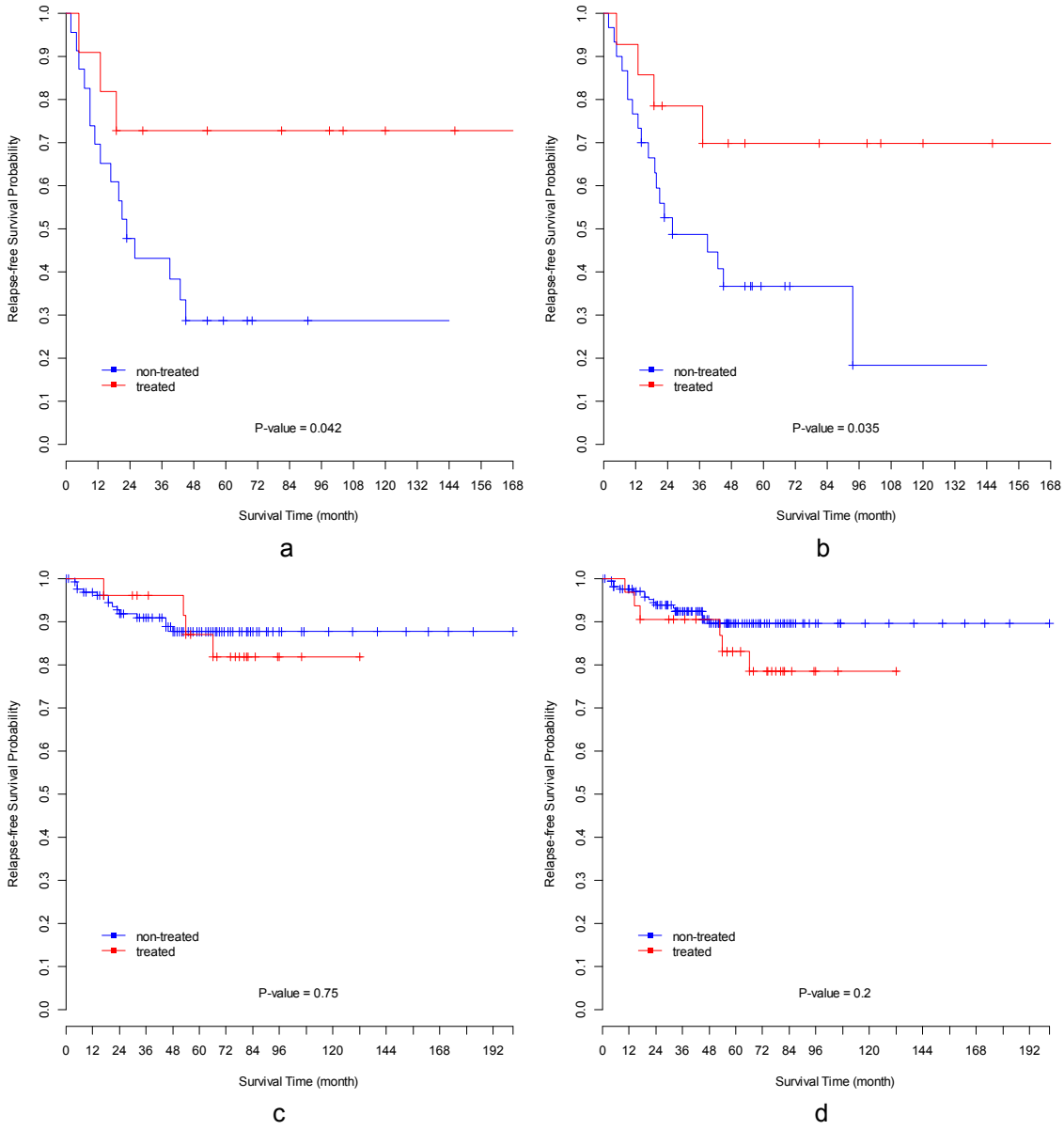
Notes:
*Percentage of non-recur (i.e., 'real low-risk') samples in the predicted low-risk group.
†Percentage of the predicted low-risk samples from the non-recur group.
**Percentage of recur (i.e., 'real high-risk') samples in the predicted high-risk group.
†† Percentage of the predicted high-risk samples from the recur group.
NA: If recur samples are less than six in a dataset, the accuracy and recall rates of the high-risk group will be not calculated.
If a dataset contains less than 30 samples, it was meta-normalized with another dataset, however, both datasets must share the same microarray platform.
If a sample has no recurrence/metastasis information, we excluded it from our analysis.

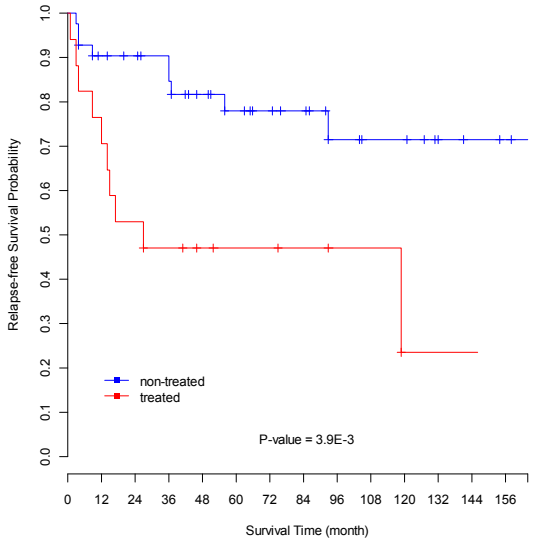
eFigure 1 Kaplan–Meier curves of the risk groups for stage II CRC patients with 5-year disease-free survival predicted by the CSS sets



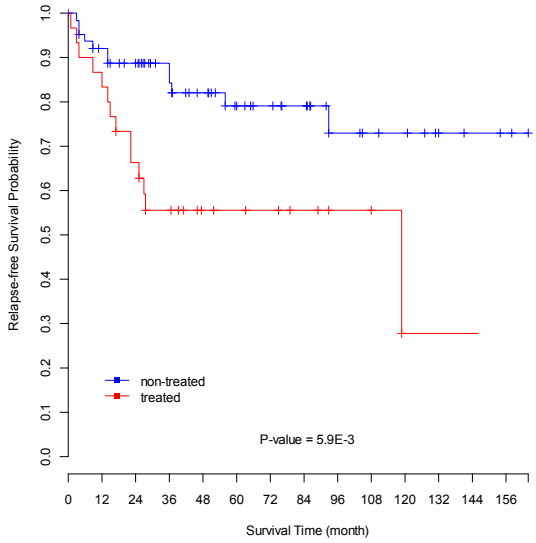
(**a**) GSE33113, (**b**) GSE14333 and (**c**) Metadata of GSE16125 and GSE24551. Drug treated samples in these cohorts have been excluded from the analysis. Kaplan–Meier curves have been not made for the cohorts which contain only metastasis information but without follow-up time. Black, blue and red curves represent low- intermediate- and high-risk groups, respectively. P-values were obtained from the χ^2-test.

eFigure 2 Kaplan–Meier plot and linear fit plot for distance recurrences comparing stage II CRC patients treated vs non-treated 5-FU

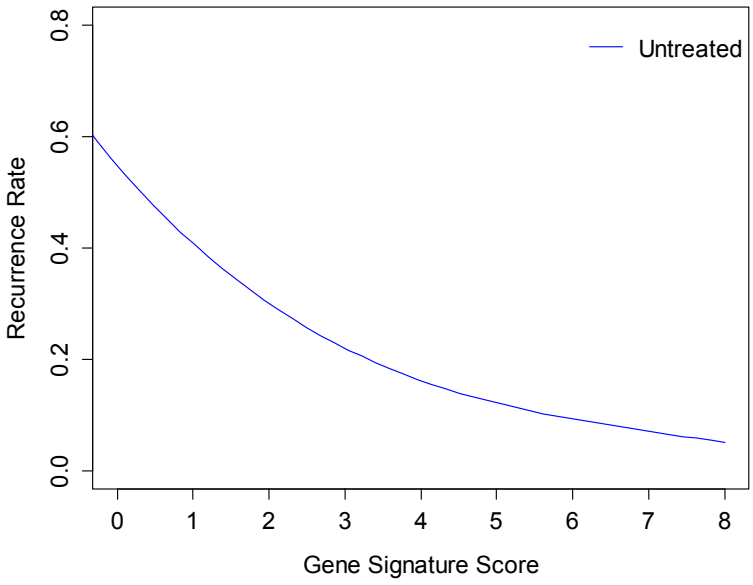




e



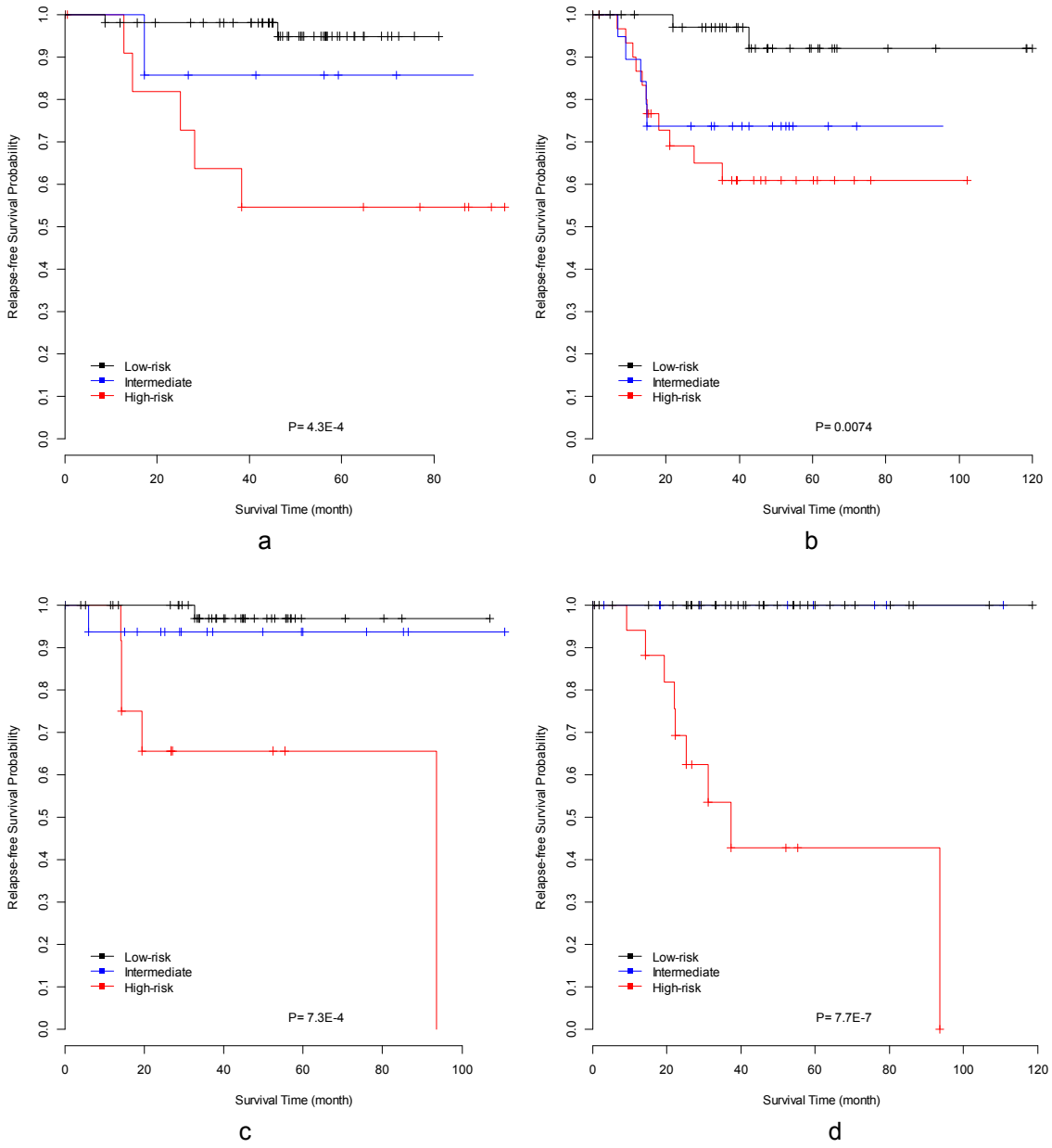
f

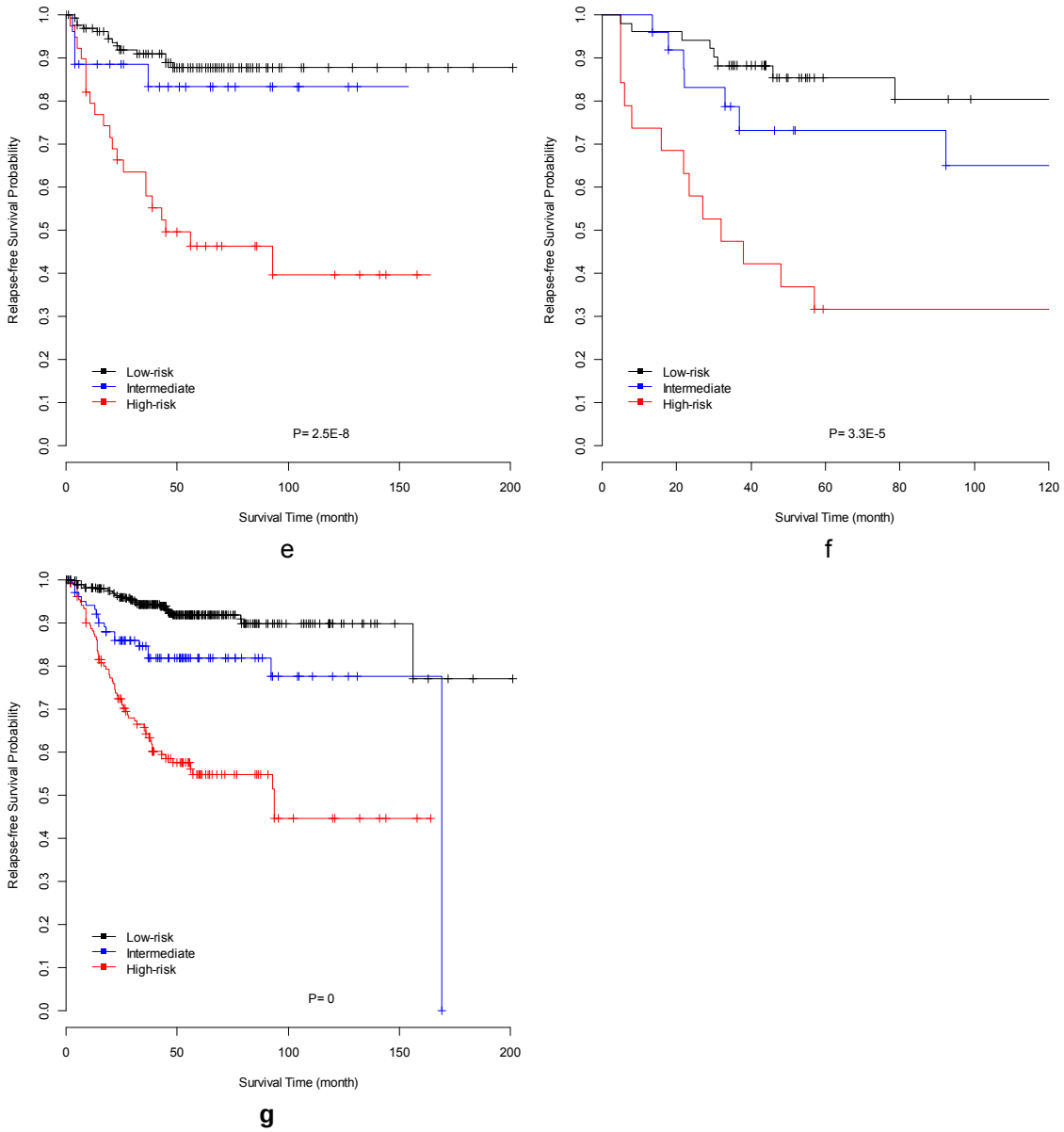


g

Kaplan–Meier plot for the CSS sets-defined ‘high-risk’ stage II patients from (**a**) GSE39582 and (**b**) GSE39582 and GSE14333. CSS sets-defined ‘low-risk’ stage II patients from (**c**) GSE39582 and (**d**) GSE39582 and GSE14333. CSS sets-defined ‘intermediate-risk’ stage II patients from (**e**) GSE39582 and (**f**) GSE39582 and GSE14333. Linear fit of the likelihood of recurrence as a continuous function of gene signature score for (**g**) all the non-treated samples. Red and blue curves represent treated and non-treated groups, respectively. P-values were obtained from the χ^2-test.

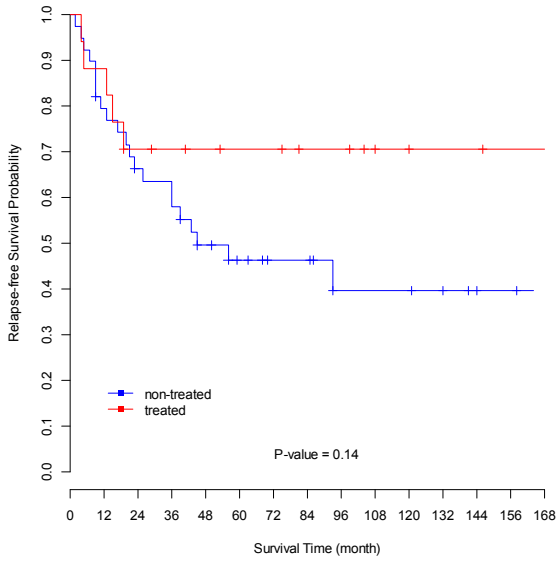
eFigure 3 Kaplan–Meier plot of the risk groups for stage II CRC patients with 5-year disease-free survival predicted by the CSS sets (the CSS set constructing rules derived from eTables 23-24)



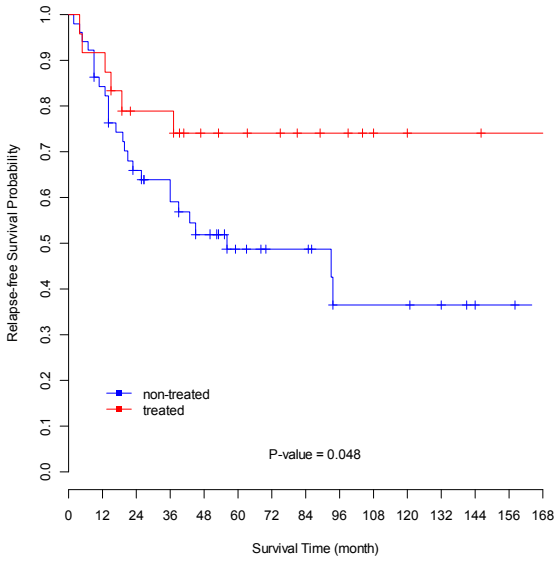


(**a**) GSE37892, (**b**) GSE33113, (**c**) GSE14333, (**d**) GSE17538, (**e**) GSE39582, (**f**) Metadata of GSE16125 and GSE24551 and (**g**) all the samples merged from above cohorts. Drug treated samples in these cohorts have been excluded from the analysis. Kaplan–Meier curves have been not made for the cohorts which contain only metastasis information but without follow-up time. Black, blue and red curves represent low- intermediate- and high-risk groups, respectively. P-values were obtained from the χ^2-test.

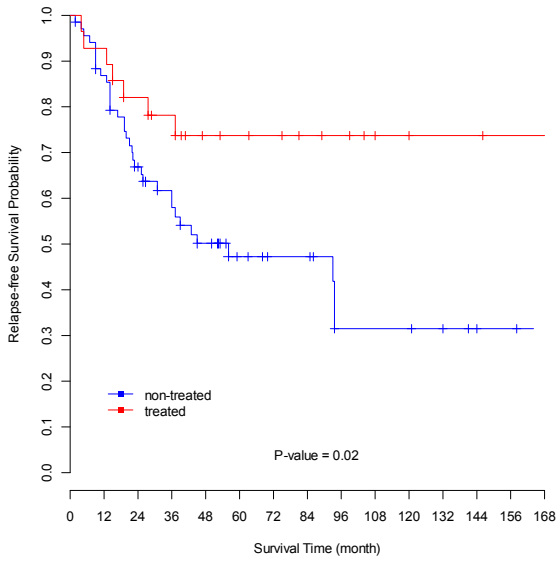
eFigure 4 Kaplan–Meier plot for distance recurrences comparing stage II CRC patients treated vs non-treated with 5-FU (the CSS sets derived from eTables 23-24)



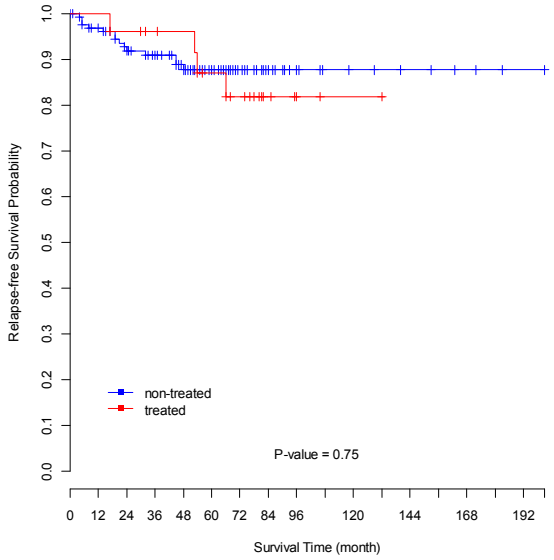
a



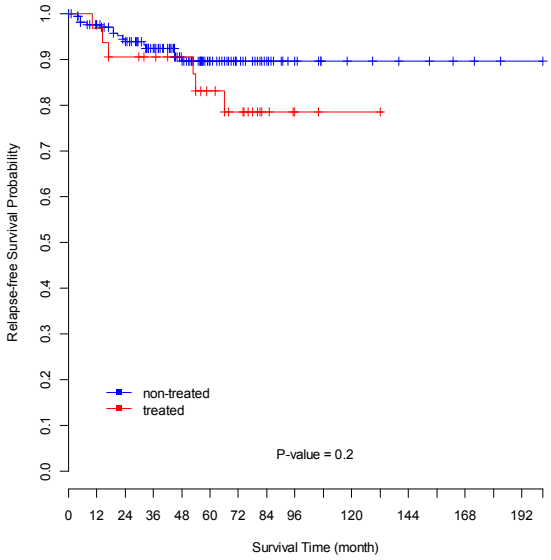
b



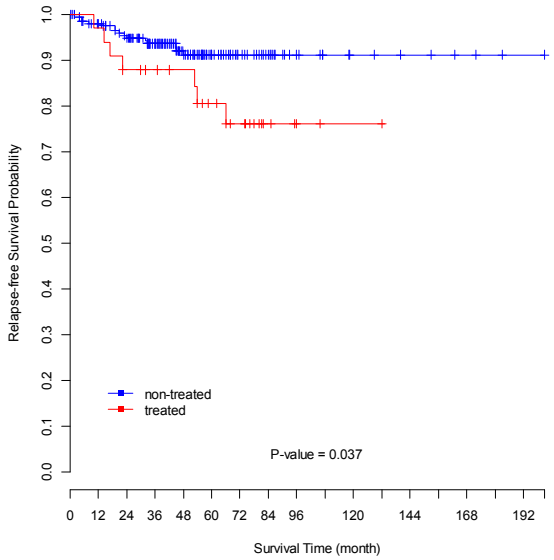
c



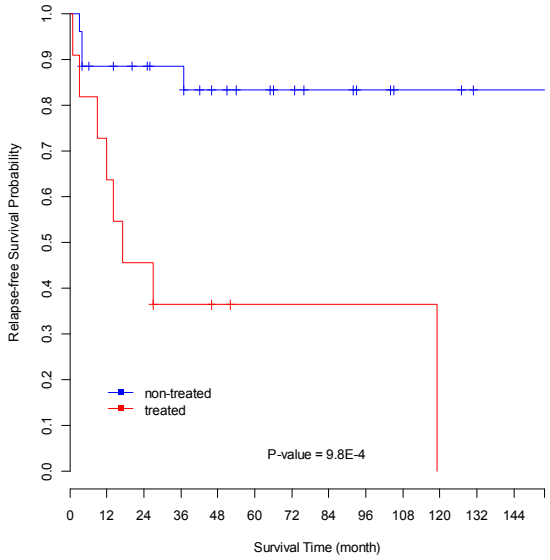
d



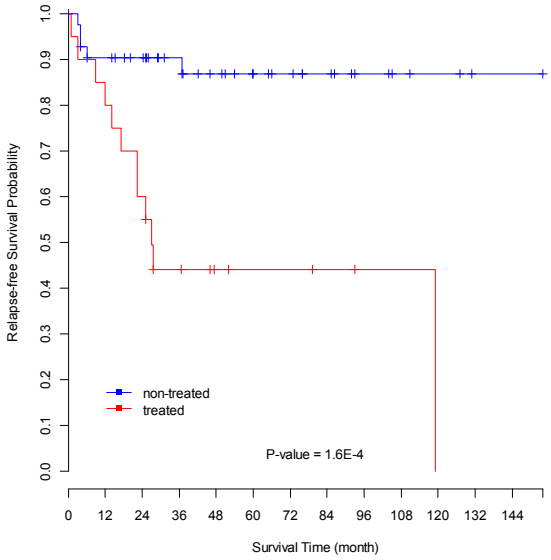
e



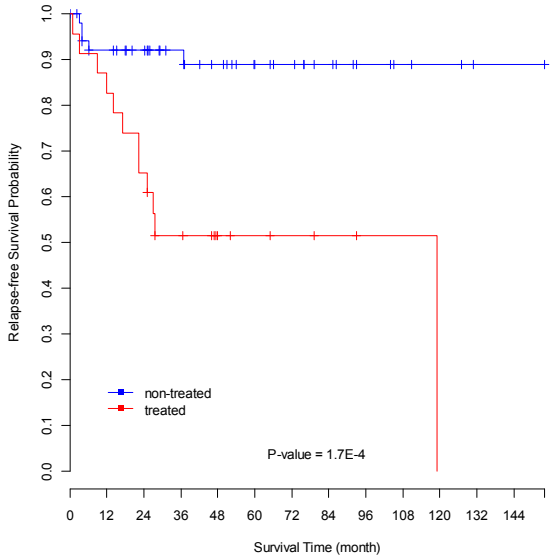
f



g



h



i

The CSS sets-defined ‘high-risk’ stage II patients from (**a**) GSE39582. (**b**) GSE39582 and GSE14333 and (**c**) GSE39582, GSE14333 and GSE17538. The CSS sets-defined ‘low-risk’ stage II patients from (**d**) GSE39582. (**e**) GSE39582 and GSE14333 and (**f**) GSE39582, GSE14333 and GSE17538. The CSS sets-defined ‘intermediate-risk’ stage II patients from (**g**) GSE39582. (**h**) GSE39582 and GSE14333 and (**i**) GSE39582, GSE14333 and GSE17538. Red and blue curves represent treated and non-treated groups, respectively. P-values were obtained from the χ^2-test.

eReferences

(1) Cui Q, Ma Y, Jaramillo M et al. A map of human cancer signaling. *Mol Syst Biol* 2007;3:152.

(2) Dennis G, Jr., Sherman BT, Hosack DA et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4:3.

(3) Li J, Lenferink AE, Deng Y et al. Identification of high-quality cancer prognostic markers and metastasis network modules. *Nat Commun* 2010;1:34.

(4) Park YY, Lee SS, Lim JY et al. Comparison of prognostic genomic predictors in colorectal cancer. *PLoS One* 2013;8:e60778.

(5) Wang Y, Jatkoe T, Zhang Y et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. *J Clin Oncol* 2004;22:1564-1571.

(6) Jorissen RN, Gibbs P, Christie M et al. Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. *Clin Cancer Res* 2009;15:7642-7651.

(7) Agesen TH, Sveen A, Merok MA et al. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. *Gut* 2012;61:1560-1567.

(8) Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A* 2002;99:6567-6572.

(9) Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646-674.

(10) Wang E, Zaman N, McGee S, Milanese JS, Masoudi-Nejad A, O'Connor-McCourt M. Predictive genomics: a cancer hallmark network framework for predicting tumor clinical phenotypes using genome sequencing data. *Semin Cancer Biol* 2015;30:4-12.