Claims

[Claim 1]. A method for predicting the treatment response to chemotherapy of a human gastroesophageal cancer patient, the method comprising:

a) measuring the gene expression of at least the genes: CDH1, CDK6, COX2, ELOVL5, GATA4, EGFR, TBCEL, FGF7, CDH17, FNBP1, PIP5K1B, TWIST, CD44, MET, CEACAM1, TOX3, GLIPR2, GSTP1, RON, TMEM136, MYB, BRCA2, FGF1, POU5F1, EPR, DPYD, ABL2 and SH3RF1 in a sample obtained from the gastroesophageal tumour of the patient to obtain a sample gene expression profile of at least said genes CDH1,ELOVL5, EGFR, PIP5K1B, FGF1, CD44 and TBCEL; and

b) making a prediction of the treatment response and/or prognosis of the patient based on the sample gene expression profile.

[Claim 2]. The method of claim 1, wherein the patient is a patient who has had perioperative chemotherapy and surgical resection of the gastroesophageal tumour.

[claim 3]. The method of claim 1, wherein step b) making a prediction of the treatment response of the patient based on the sample gene expression profile comprises:

(i) optionally, normalising the measured expression level of each gene relative to the expression level of one or more housekeeping genes;

(ii) computing a risk score by weighting the measured, and optionally normalised, expression level of each gene and summing the weighted expression level of each of the genes, wherein the contribution to the total risk score made by CD44 and EGFR has the opposite sign to that of the contribution made by CDH1, ELOVL5, PIP5K1B, FGF1 and TBCEL.

[Claim 4]. The method of claim 3, wherein the risk score is referenced to the median risk score of a sample cohort of gastric cancer patients, which median risk score serves as a threshold, and wherein:

a computed risk score of above that threshold indicates that the patient is at high risk of poor treatment response, at high risk of suffering recurrence of the tumour and/or at high risk of having a shorter than median survival time; and

a computed risk score of below that threshold indicates that the patient is at low risk of poor treatment response, at low risk of suffering recurrence of the tumour and/or at low risk of having a shorter than median survival time.

[Claim 5]. The method of claim 3 or claim 4, wherein the risk score for the patient is calculated by taking, for each gene, the product of the hazard ratio (HR) for that gene and the measured, and optionally normalised, gene expression value, summing those products for all seven genes, wherein the sign of the genes CDH1, ELOVL5, PIP5K1B, FGF1 and TBCEL is negative, lowering the total risk score with increasing expression, and the sign of the genes CD44 and EGFR is positive, increasing the total risk score with increasing expression.

[Claim 6]. The method of claim 1, wherein step b) making a prediction of the treatment response of the patient based on the sample gene expression profile comprises:

(i) optionally, normalising the measured expression level of each gene relative to the expression level of one or more housekeeping genes;

(ii) comparing the sample gene expression profile, optionally after said normalising, with two or more reference centroids comprising:

a first reference centroid that represents the average gene expression of at least the genes CDH1, ELOVL5, EGFR, PIP5K1B, FGF1, CD44 and TBCEL measured in a low risk training set made up of gastroesophageal cancer patients known to have had a median survival time following tumour resection of greater than 3 years or greater than 6 years; and

a second reference centroid that represents the average gene expression of at least the genes CDH1, ELOVL5, EGFR, PIP5K1B, FGF1, CD44 and TBCEL measured in a high risk training set made up of gastroesophageal cancer patients known to have had a median survival time following tumour resection of less than 1.5 years or less than 1 year;

c) classifying the sample gene expression profile as belonging to the risk group having the reference centroid to which it is most closely matched; and

d) providing a prediction of treatment response or prognosis based on the classification made in step c).

[Claim 7]. The method of claim 1, wherein step b) making a prediction of the treatment response of the patient based on the sample gene expression profile comprises:

(i) optionally, normalising the measured expression level of each gene relative to the expression level of one or more housekeeping genes;

(ii) comparing the sample gene expression profile, optionally after said normalising, with at least three reference centroids corresponding to low,

moderate and high risk subgroups, respectively, the reference centroid comprising:

a first reference centroid that represents the average gene expression of at least the genes CDH1, ELOVL5, EGFR, PIP5K1B, FGF1, CD44 and TBCEL measured in a low risk training set made up of gastroesophageal cancer patients known to have had median survival time following tumour resection of greater than 7 years;

a second reference centroid that represents the average gene expression of at least the genes CDH1, ELOVL5, EGFR, PIP5K1B, FGF1, CD44 and TBCEL measured in a moderate risk training set made up of gastroesophageal cancer patients known to have had median survival time following tumour resection of between 1.4 and 4.5 years; and

a third reference centroid that represents the average gene expression of at least the genes CDH1, ELOVL5, EGFR, PIP5K1B, FGF1, CD44 and TBCEL measured in a high risk training set made up of gastroesophageal cancer patients known to have had median survival time following tumour resection of less than 0.6 years or even less than 0.4 years;

c) classifying the sample gene expression profile as belonging to the risk group having the reference centroid to which it is most closely matched; and

d) providing a prediction of treatment response or prognosis based on the classification made in step c).

[Claim 8]. The method of any one of the preceding claims, wherein the gene expression profile is of an expanded gene set comprising at least the genes CDH1, ELOVL5, EGFR, PIP5K1B, FGF1, CD44, TBCEL, FGF7, CDK6, GLIPR2, FNBP1, TOX3, ABL2, RON, CDH17, GATA4, TWIST, COX2, BRCA2, DPYD, CEACAM1, EPR, MET, TMEM136, MYB, SH3RF1, POU5F1 and GSTP1.

[Claim 9]. The method of claim 8, wherein step b) making a prediction of the treatment response of the patient based on the sample gene expression profile comprises:

(i) optionally, normalising the measured expression level of each gene relative to the expression level of one or more housekeeping genes;

(ii) comparing the sample gene expression profile, optionally after said normalising, with at least two reference centroids corresponding to low and high risk subgroups, respectively, the reference centroids comprising:

Genes	Low-Risk	High-Risk
ELOVL5	0.1967	-0.1967
TBCEL	0.1961	-0.1961
CDH1_201131_s_at	0.1903	-0.1903
FGF7	0.1737	-0.1737
CD44v8.10	-0.1656	0.1656
PIP5K1B	0.1538	-0.1538
CDK6	0.1513	-0.1513
GLIPR2	0.1379	-0.1379
FNBP1	0.1308	-0.1308
EGFR	-0.1299	0.1299
TOX3	0.1212	-0.1212
ABL2	0.1179	-0.1179
CDH1_201130_s_at	0.1044	-0.1044
FGF1	0.1033	-0.1033
RON	-0.1003	0.1003
CDH17	0.0996	-0.0996
GATA4	-0.0929	0.0929
TWIST	0.0848	-0.0848
COX2	0.0739	-0.0739
BRCA2	0.0664	-0.0664
DPYD	0.0635	-0.0635
CEACAM1	0.0607	-0.0607
EPR	0.0568	-0.0568
MET	-0.0452	0.0452
TMEM136	0.043	-0.043
MYB	0.0206	-0.0206
SH3RF1	0.0129	-0.0129
POU5F1	0.0127	-0.0127
GSTP1	0.0043	-0.0043

.

[Claim 10]. The method of claim 8, wherein step b) making a prediction of the treatment response of the patient based on the sample gene expression profile comprises:

(i) optionally, normalising the measured expression level of each gene relative to the expression level of one or more housekeeping genes;

(ii) comparing the sample gene expression profile, optionally after said normalising, with at least three reference centroids corresponding to low,

moderate and high risk subgroups, respectively, the reference centroids comprising:

Genes	Low	Moderate	High
TBCEL	0.2631	0.0682	-0.3694
EGFR	-0.1285	-0.1241	0.3536
FGF7	0.1373	0.119	-0.3519
CDH17	0.0807	0.1335	-0.3282
ELOVL5	0.2348	0.0564	-0.3212
CDH1_201131_s_at	0.3146	-0.0041	-0.2782
TWIST	0.0257	0.1469	-0.3039
CD44v8.10	-0.2954	0.0998	0.078
ABL2	0.2853	-0.0256	-0.2104
FGF1	0.283	-0.0104	-0.2375
CDH1_201130_s_at	0.2752	-0.034	-0.1852
FNBP1	0.1321	0.0756	-0.2645
CDK6	0.2446	-0.0149	-0.194
PIP5K1B	0.2446	0.0037	-0.2295
TOX3	0.1031	0.0698	-0.227
GLIPR2	0.1334	0.0462	-0.2095
GATA4	0.0048	-0.1038	0.1937
COX2	0.1143	0.0466	-0.1928
BRCA2	0.0496	0.0749	-0.188
MET	-0.1698	0.0384	0.0811
POU5F1	0.1662	-0.0272	-0.0991
TMEM136	0.1508	-0.0421	-0.0566
MYB	0.0064	0.0635	-0.1271
CEACAM1	0.0886	0.0147	-0.1086
GSTP1	0.0172	0.0436	-0.0989
RON	-0.0625	-0.0208	0.0965
DPYD	0.0506	-0.0154	-0.0166
EPR	-0.0284	0.0345	-0.04
SH3RF1	0.0297	-0.0336	0.037

.

[Claim 11]. The method of claim 1, wherein the gene expression profile is of an expanded gene set comprising at least the genes CDH1, CDK6, COX2, ELOVL5, GATA4, EGFR,

TBCEL, FGF7, CDH17, FNBP1, PIP5K1B, TWIST, CD44MET, CEACAM1, TOX3, GLIPR2, GSTP1, RON, TMEM136, MYB, CDH1, BRCA2, FGF1, POU5F1, EPR, DPYD and SH3RF1.

[Claim 12]. The method of claim 11, wherein step b) making a prediction of the treatment response of the patient based on the sample gene expression profile comprises:

(i) optionally, normalising the measured expression level of each gene relative to the expression level of one or more housekeeping genes;

(ii) comparing the sample gene expression profile, optionally after said normalising, with at least three reference centroids corresponding to low, moderate and high risk subgroups, respectively, the reference centroid comprising:

	Centroids		
Genes	1. High Risk	2. Low Risk	3. Moderate Risk
FGF7	-1.5785	0.0235	0.2413
CDH17	-1.2605	-0.162	0.2794
FNBP1	-1.0724	0.1664	0.0917
TBCEL	-0.9271	0.2638	0.0217
PIP5K1B	-0.8545	0.1563	0.0617
ELOVL5	-0.7801	0.2959	-0.0172
TWIST	-0.7556	-0.021	0.131
CD44v8.10	0.0687	-0.6475	0.2998
MET	0.628	-0.4301	0.106
COX2	-0.6045	0.0574	0.0692
CEACAM1	-0.5618	0.1421	0.0217
TOX3	-0.5553	0.0584	0.0608
GLIPR2	-0.5525	0.1912	-0.0034
GSTP1	-0.4809	0.0578	0.0492
RON	0.4109	-0.309	0.0826
TMEM136	-0.3886	0.2811	-0.0727
MYB	-0.3709	-0.0939	0.1044
CDH1_201130_s_at	-0.128	0.3616	-0.1531
BRCA2	-0.3143	-0.0752	0.0864
CDH1_201131_s_at	-0.1409	0.2648	-0.1046
FGF1	-0.2643	0.1278	-0.019
CDK6	0.0886	0.2509	-0.1346
POU5F1.	-0.2326	0.2432	-0.0795
EPR	0.028	-0.2399	0.1107
DPYD	0.1492	0.0302	-0.0384
SH3RF1	0.1005	1.00E-04	-0.0161
EGFR	0.0974	-0.0777	0.0217
GATA4	0.0698	0.067	-0.0433

.

[Claim 13]. The method of any one of the preceding claims, wherein the patient, having been determined to be at high or moderate risk of poor treatment response or poor prognosis, is selected for additional or alternative treatment, including aggressive treatment.

[Claim 14]. The method of any one of the preceding claims, wherein the patient, having been found to be at low risk of poor treatment response or low risk of poor prognosis, is selected less aggressive ongoing treatment or for non-treatment.

[Claim 15]. A computer-implemented method for predicting the treatment response to chemotherapy or prognosis of a human gastroesophageal cancer patient, the method comprising:

a) obtaining gene expression data comprising a gene expression profile representing gene expression measurements of at least the genes CDH1, ELOVL5, EGFR, PIP5K1B, FGF1, CD44 and TBCEL measured in a sample obtained from the gastroesophageal tumour of the patient; and

b) (i) optionally, normalising the measured expression level of each gene relative to the expression level of one or more housekeeping genes,

(ii) comparing the sample gene expression profile with two or more reference centroids as defined in table 16 or table 18;

c) classifying the sample gene expression profile as belonging to the risk group having the reference centroid to which it is most closely matched; and

d) providing a prediction of treatment response or prognosis based on the classification made in step c).