Supplementary Materials

Supplementary Table 1. Non-linear mixed effect model parameter estimation.

Parameter	Definition	Unit	Estimate	Standard Error	Relative Standard Error (%)
Kg	Tumor progression rate	1/day	0.0033	5.04E-5	1.53
Kd	Tumor regression rate	1/day	0.0171	0.000152	0.887
F	Resistant cell fraction	1	0.121	0.00306	2.53
Ωkg	Inter-lesion variability of Kg (log-normal distribution)		1.24	0.0117	0.941
Ωkd	Inter-lesion variability of Kd (log-normal distribution)		0.732	0.00771	1.05
ΩF	Inter-lesion variability of F (logit distribution)		2.91	0.0248	0.852
b	Proportional error model		0.296	0.000904	0.305

Supplementary Table 2. Demographic information of head and neck squamous cell carcinomas patients in validation dataset.

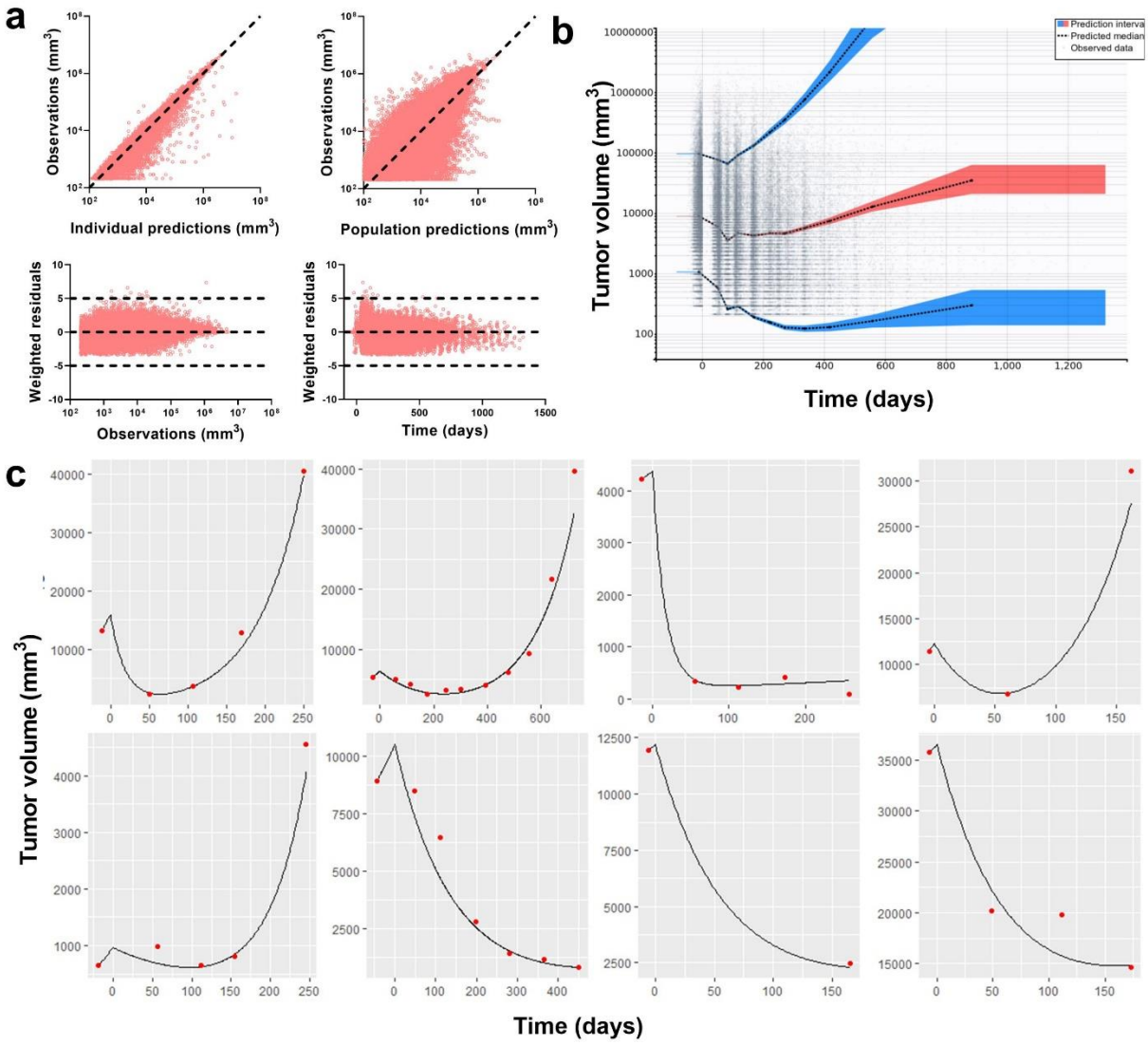
Variable	N=393
Age, years (mean, sd)	57.8 (7.9)
Gender (n, %)	
Male	346 (88.0)
Female	47 (12.0)
Race (n, %)	
White/Caucasian	356 (90.6)
Asian	34 (8.7)
Hispanic/Latino	1 (0.3)
Other	2 (0.4)
Body Surface Area, m^2 (mean,sd)	1.8 (0.2)
Tumor Type (n, %)	
Hypopharynx	60 (15.3)
Larynx	124 (31.6)
Oral Cavity	104 (26.4)
Oropharynx	105 (26.7)
Prior Surgery (n, %)	
Yes	334 (85.0)
No	59 (15.0)
Prior Radiation (n, %)	
Yes	305 (77.6)
No	88 (22.4)
Treatment (n, %)	
Panitumumab plus Chemotherapy	198 (50.4)
Chemotherapy Alone	195 (49.6)
Response (n, %)	
Complete Response	11 (2.8)
Partial Response	122 (31.0)
Progressive Disease	42 (10.7)
Stable Disease	215 (54.7)
Not Evaluable	3 (0.8)
Metastatic organ number (n, %)	
1	122 (31.0)
2	145 (36.9)
3	87 (22.1)
>=4	39 (9.9)

Supplementary Table 3. Demographic and metastatic profiles of patients under k-means clusters.

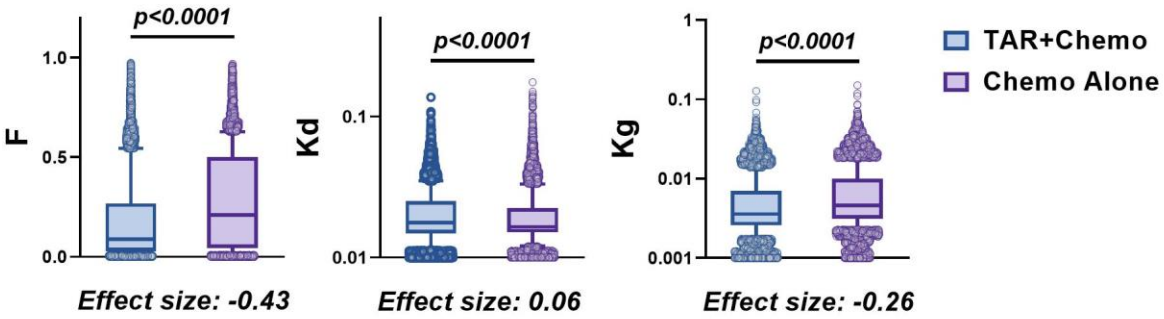
Variable	Hetero-Organ (n=817)	Liver-First (n=930)	Lung-First (n=577)	Mono-Organ (n=1,345)	Other-First (n=639)
Age, years (mean, sd)	59.8 (11.1)	60.5 (10.7)	61.5 (10.5)	60.1 (10.6)	59.8 (11.2)
Gender (n, %)					
Male	476 (58.3)	562 (60.4)	351 (60.8)	793 (59.0)	356 (55.7)
Female	341 (41.7)	368 (39.6)	226 (39.2)	552 (41.0)	283 (44.3)
Race (n, %)					
White/Caucasian	754 (92.3)	844 (90.8)	521 (90.3)	1,205 (89.6)	559 (87.5)
Black/African American	13 (1.6)	19 (2.0)	12 (2.1)	41 (3.1)	19 (3.0)
Asian	17 (2.1)	31 (3.3)	19 (3.3)	49 (3.6)	26 (4.0)
Other	33 (4.0)	36 (3.9)	25 (4.3)	50 (3.7)	35 (5.5)
BMI, kg/m^2 (mean,sd)	26.1 (4.9)	26.1 (5.0)	26.5 (5.1)	26.2 (5.0)	26.4 (5.1)
Tumor Type (n, %)					
Colon	488 (59.7)	550 (59.1)	281 (48.7)	866 (64.4)	396 (62.0)
Rectal	265 (32.4)	303 (32.6)	256 (44.4)	356 (26.5)	179 (28.0)
Unspecified	64 (7.8)	77 (8.3)	40 (6.9)	123 (9.1)	64 (10.0)
Prior surgery (n, %)					
Yes	565 (69.2)	628 (67.5)	379 (65.7)	969 (72.0)	452 (70.7)
No	252 (30.8)	302 (32.5)	198 (34.3)	376 (28.0)	187 (29.3)
Prior radiation (n, %)					
Yes	588 (72.0)	705 (75.8)	390 (67.6)	1,160 (86.2)	502 (78.6)
No	89 (10.9)	88 (9.5)	101 (17.5)	91 (6.8)	76 (11.9)
Unknown	140 (17.1)	137 (14.7)	86 (14.9)	94 (7.0)	61 (9.5)
Metastatic organ number (n, %)					
1	0 (0)	0 (0)	53 (9.2)	482 (35.8)	18 (2.8)
2	0 (0)	305 (32.8)	185 (32.1)	444 (33.0)	225 (35.2)
3	176 (21.5)	315 (33.9)	188 (32.6)	270 (20.1)	197 (30.8)
≥4	641 (78.5)	310 (33.3)	151 (26.2)	149 (11.1)	199 (31.1)
KRAS status (n, %)					
Wild-Type	129 (15.8)	172 (18.5)	88 (15.3)	272 (20.2)	134 (21.0)
Mutant	106 (13.0)	142 (15.3)	75 (13.0)	190 (14.1)	80 (12.5)
Unknown	582 (71.2)	616 (66.2)	414 (71.8)	883 (65.7)	425 (66.5)
Has liver metastases at baseline (n, %)					
Yes	667 (81.6)	920 (98.9)	384 (66.6)	1,285 (95.5)	406 (63.5)
No	150 (18.4)	10 (1.1)	193 (33.4)	60 (4.5)	233 (36.5)
Baseline target tumor volume mm^3 (mean, sd)	243,091 (402,762)	240,895 (432,291)	169,061 (380,487)	192,247 (353,082)	122,878 (249,221)

Supplementary Table 4. The hyperparameters of the gradient boosting model to predict patient relapse sequence.

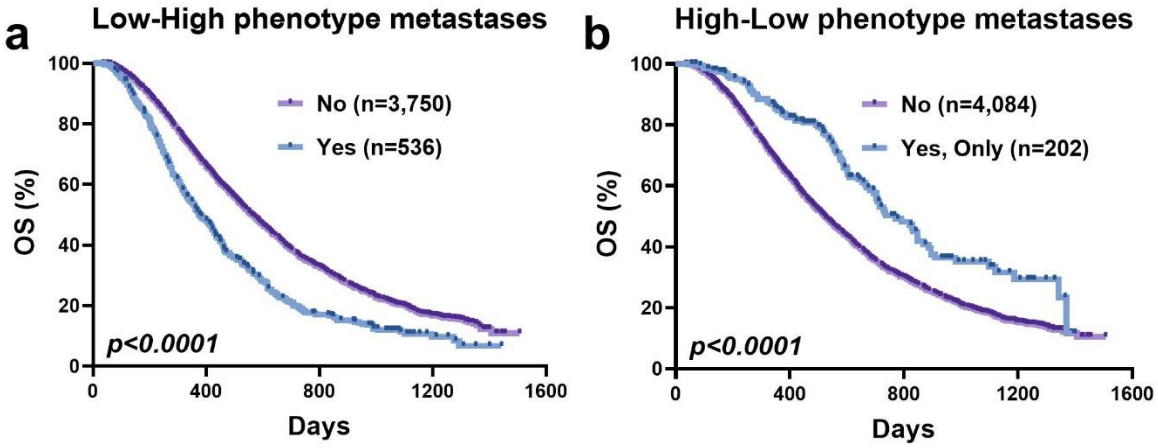
Hyperparameters	Grid-search Range	Best Performed Model
Max_depth	3, 4, 5	3
Min_child_weight	1, 2, 5	2
Learning_rate	0.1	0.1
subsample	0.5	0.5
N_estimators	50, 100, 200	200
colsample_bytree	0.5	0.5



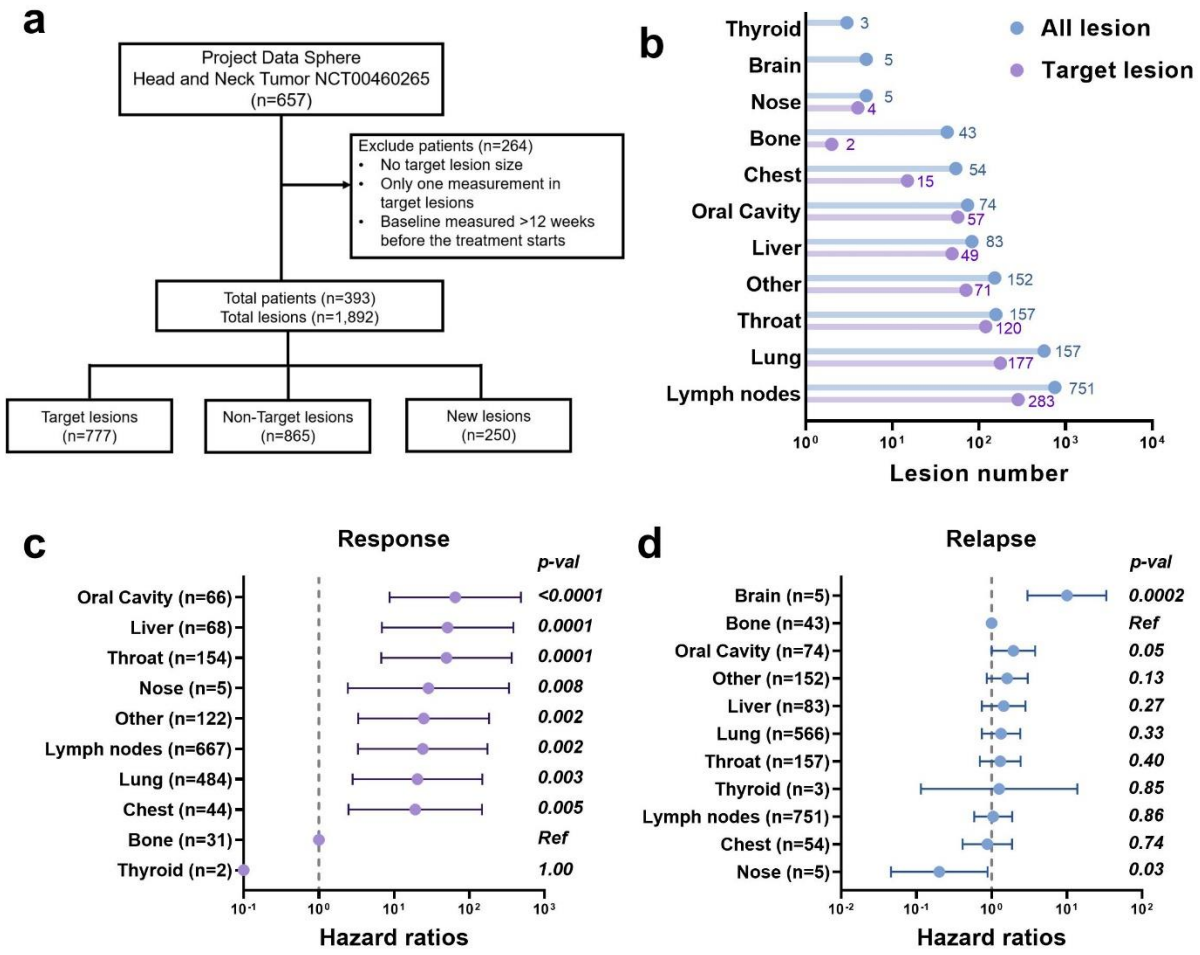
Supplementary Fig. 1 Tumor growth model has good model performance. a. Goodness-of-fit plots for tumor growth model. The upper panels are the observed tumor volume data plotted against individual model predictions (left) or population model predictions (right). The bottom panels are the weighted residuals plotted against observed tumor volume (left) or time (right). **b.** Visual predictive checks for tumor volume simulated from tumor growth model. The gray dots are the observations. The dashed lines describe the 10^{th}, 50^{th}, and 90^{th} percentiles of model predictions and the bands are the model-simulated 90% confidence interval of the corresponding percentile. **c.** Eight representative lesions observed tumor volume (red circles) and the model predictions (black line) change with time.



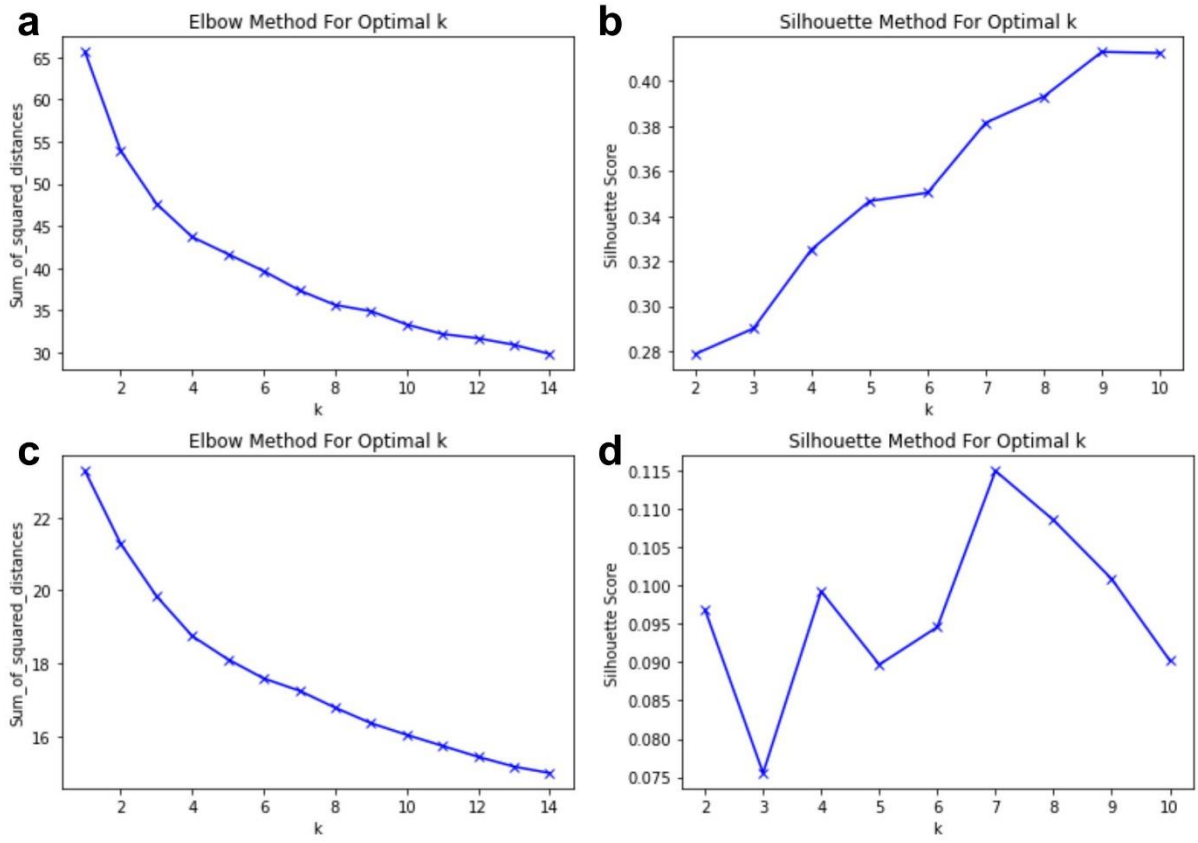
Supplementary Fig. 2 Comparisons of parameter *F*, *Kd*, *Kg* of lesions under antibody targeted therapies plus chemotherapy (TAR+Chemo) vs. chemotherapy alone (Chemo Alone). The box extends from the 25th to 75th percentiles and the line in the middle is plotted as the median. The whiskers are drawn down to the 10th percentile and up to the 90th percentile. Points below and above the whiskers are drawn as individual points. P-values are calculated by two-tailed Mann–Whitney test and the effect size are calculated by *Cohen'd*.



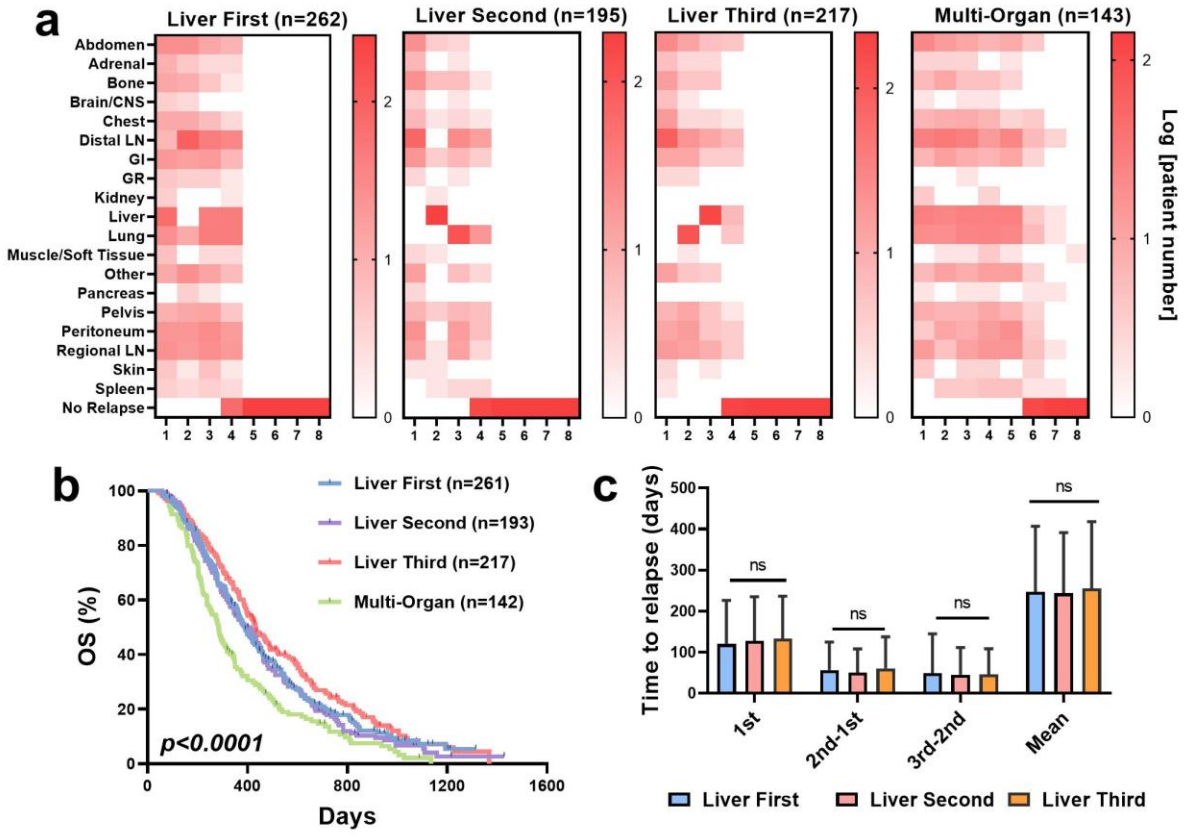
Supplementary Fig. 3 Metastasis Organ phenotypes are associated with patient prognosis. a. Patients without low-high phenotype metastases (genitourinary and reproductive system, adrenal, muscle/soft tissues, bone, brain/central nervous system) have significantly better survival than patients have low-high phenotype metastases. **b.** Patients only have metastases in high-low phenotype organs (lymph nodes, chest, spleen, lung) have significantly better survival than patients with other phenotypes.



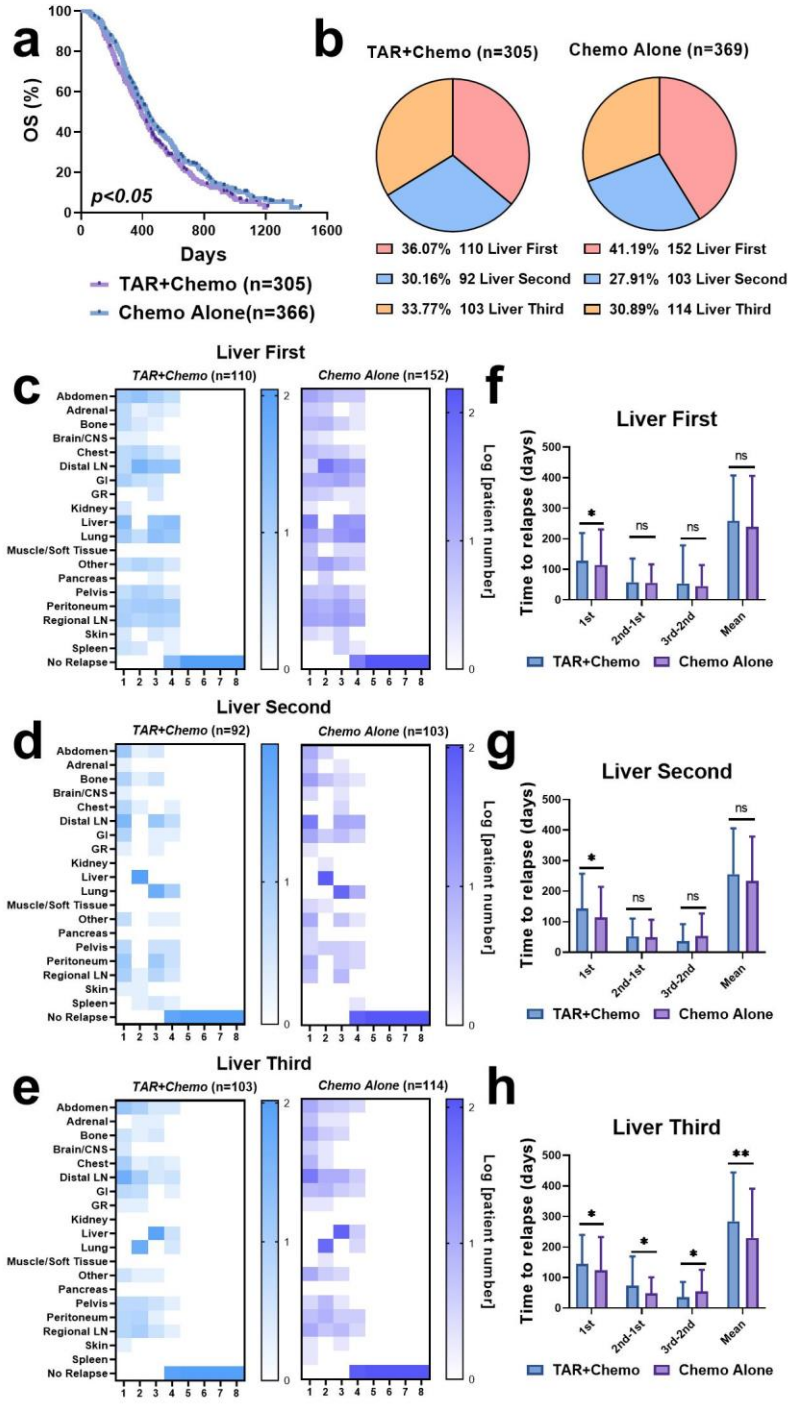
Supplementary Fig. 4 Validation of organ mapping in head and neck tumor. a. CONSORT diagram of patients and lesions included in data analysis. **b.** The number of all lesions and target lesions across organs. **c** and **d** Hazard ratios and 95% confidence interval with p-values of lesion response and relapse by organs.



Supplementary Fig. 5 Determine optimal k for K-means. **a** and **b** Elbow method and Sihouette method were applied to find optimal k in relapse patterns classification for all the patients. **c** and **d** Elbow method and Sihouette method were applied to find optimal k in relapse patterns classification for Hetero-Organ patients.

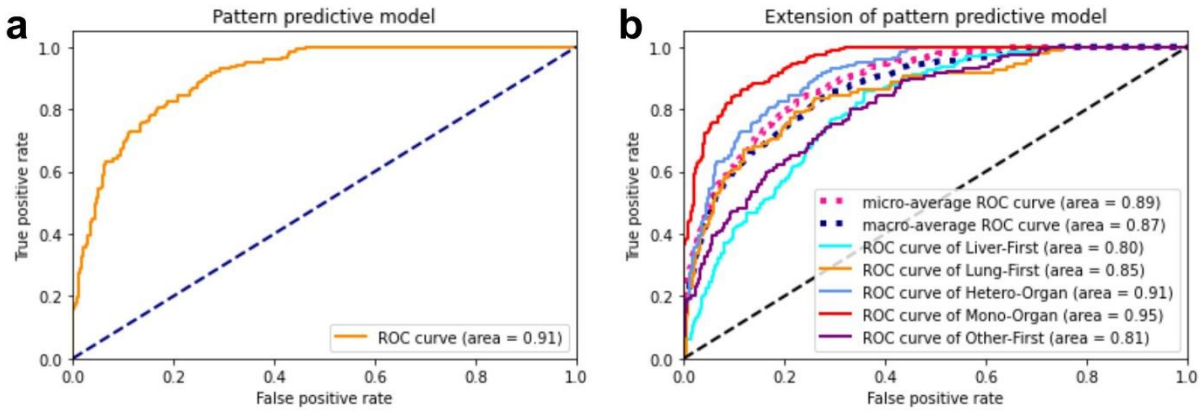


Supplementary Fig. 6 Liver relapse sequence is associated with patient survival. a. Hetero-Organ patients were clustered into four groups by liver relapse sequqnce. **b.** Kaplan-Meier curves of patient overall survival by liver relapse sequence. **c.** The mean and standard deviation of the first lesion relapse time (1st), time between first and second relapse (2nd-1st), time between second and third relapse (3rd-2nd), time between third and fourth relapse (4th-3rd), and the average relapse time in Liver First (n=261), Liver Second (n=193), Liver Third (n=217), and Multi-Organ groups (n=142). ns, not significance (p-value >0.05).



Supplementary Fig. 7 Targeted therapy has minimal effect on sequence order. a. Three hetero-organ sub-groups in **Supplementary Fig. 6a** (Liver First, Liver Second, and Liver Third) patients overall survival stratified by treatments. **b.** Liver First, Liver Second, and Liver Third patient proportions by treatments. **c, d,** and **e** Patient relapse sequences stratified by treatments. **f, g,** and **h** The mean and standard deviation of the first lesion relapse time (1st), time between first and second relapse (2nd-1st), time between second and third relapse (3rd-2nd), time between third and fourth relapse (4th-3rd), and the

average relapse time by treatments of the groups in **c, e,** and **g**. $p<0.0001$, ****; $p<0.01$, **; $p<0.05$, *; $p>0.05$, ns.



Supplementary Fig. 8 Relapse sequence prediction model performance. **a.** The machine learning algorithm Gradient Boosting (XGBoost) model overall performance. **b.** Model performance stratified by sub-cluster. The results of the training group were: precision=0.76, accuracy =0.75, recall = 0.76. The results of the testing group were: precision = 0.59, accuracy = 0.60, recall = 0.58.