Supplemental Data

Enhancing Mass spectrometry-based tumor immunopeptide identification: machine learning filter leveraging HLA binding affinity, aliphatic index and retention time deviation

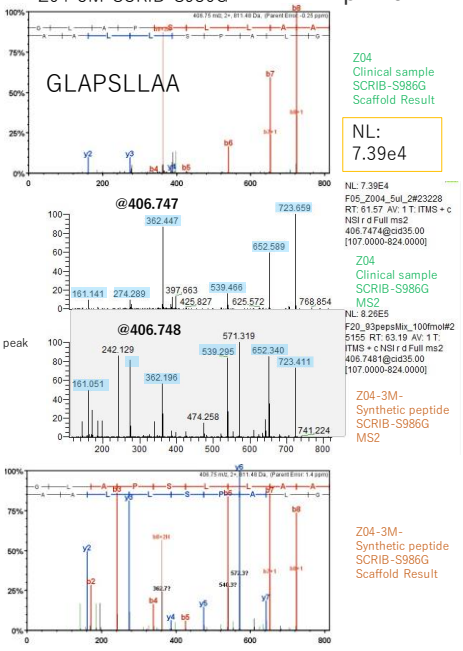
Feifei Wei, Taku Kouro, Yuko Nakamura, Hiroki Ueda, Susumu Iiizumi, Kyoko Hasegawa, Yuki Asahina, Takeshi Kishida, Soichiro Morinaga, Hidetomo Himuro, Shun Horaguchi, Kayoko Tsuji, Yasunobu Mano, Norihiro Nakamura, Takeshi Kawamura, and Tetsuro Sasada

Contents

Peptide ID: Z04-3M | Relative retention time: 1.0

Z04-3M-SCRIB-S986G Hp: 26.41

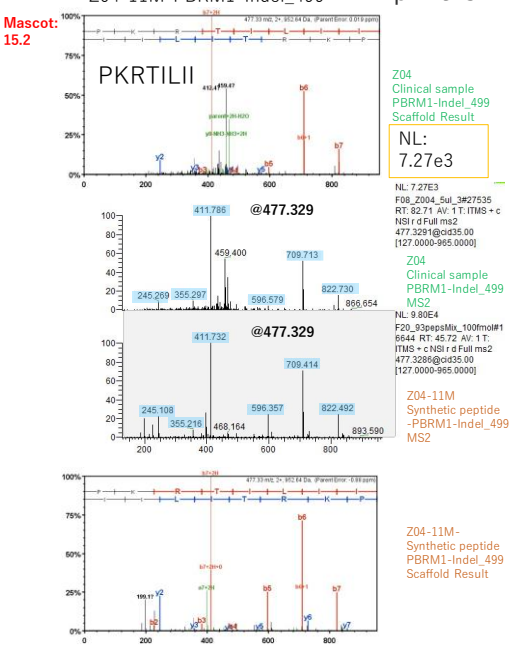
Mascot: 35.0



Peptide ID: Z04-11M | Relative retention time: 1.8

Z04-11M-PBRM1-Indel_499 Hp: 29.51

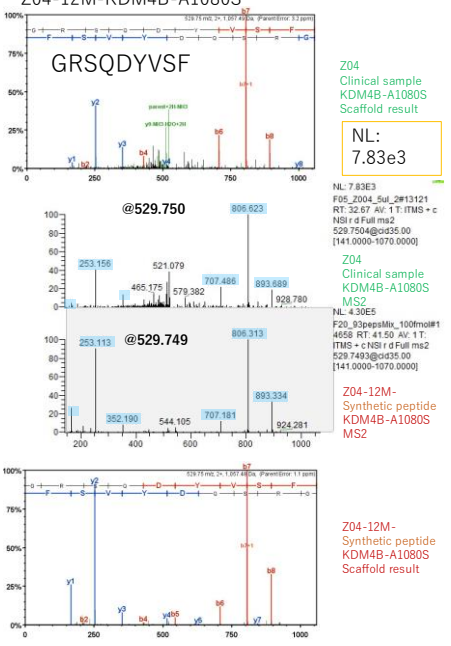
Mascot: 15.2



Peptide ID: Z04-12M | Relative retention time: 0.8

Z04-12M-KDM4B-A1080S Hp: 17.70

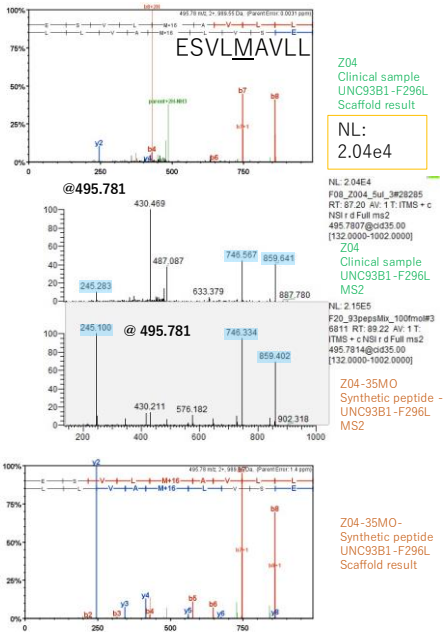
Mascot: 29.2



Peptide ID: Z04-35MO | Relative retention time: 1.0

Z04-35MO-UNC93B1-F296L Hp: 36.85

Mascot: 12.5

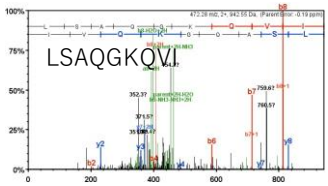


Peptide ID: Z20-16M | Relative retention time: 1.0

Z20-16M-NSMF-M299V

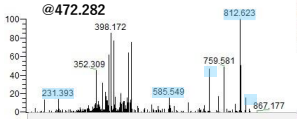
Hp: 18.34

Mascot: 11.9

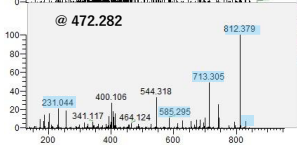


Z20 Clinical sample NSMF-M299V Scaffold result

NL: 2.58e3

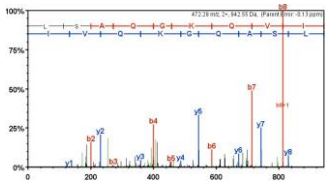


Z20Clinical sample NSMF-M299V MS2



Consistent peak

Z20-16M-Synthetic peptide NSMF-M299V MS2



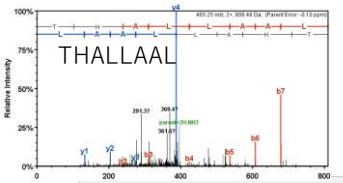
Z20-16M-Synthetic peptide NSMF-M299V Scaffold result

Peptide ID: Z19-15M | Relative retention time: 0.7

Z19-15M-SHH-A383T

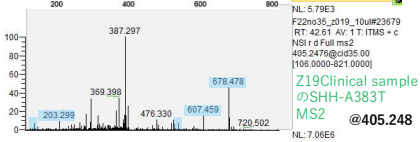
Hp: 26.60

Mascot: 11.1

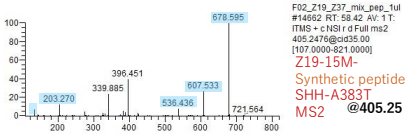


Z19 Clinical sample SHH-A383T Scaffold result

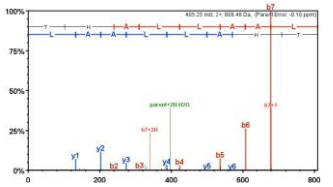
NL: 5.79e3



Z19Clinical sample of SHH-A383T MS2 @405.248



Z19-15M-Synthetic peptide SHH-A383T MS2 @405.25



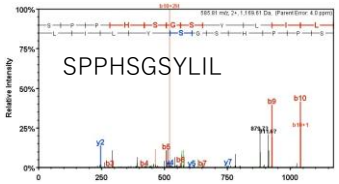
Z19-15M-Synthetic peptide SHH-A383T Scaffold result

Peptide ID: Z19-17M | Relative retention time: 1.3

Z19-17M-MRPL43-R220S

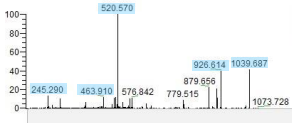
Hp: 27.91

Mascot: 12.4



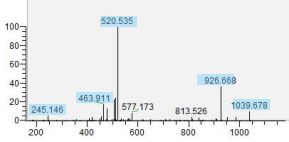
Z19 Clinical sample MRPL43-R220S Scaffold result

NL: 1.13e4

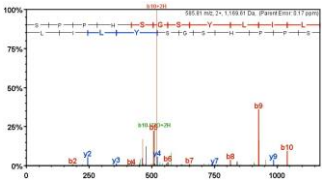


Z19Clinical sample MRPL43-R220S MS2 @585.814

Consistent peak



Z19-17M-Synthetic peptide MRPL43-R220S MS2 @585.81



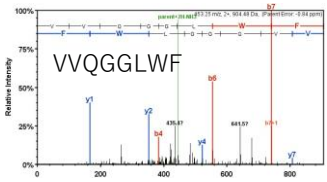
Z19-17M-Synthetic peptide MRPL43-R220S Scaffold result

Peptide ID: Z19-18M | Relative retention time: 0.1

Z19-18M-KRTAP5-1 -L174V

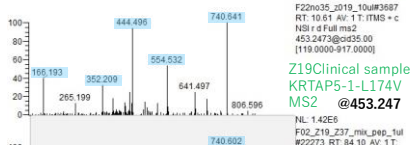
Hp: 33.15

Mascot: 13.8

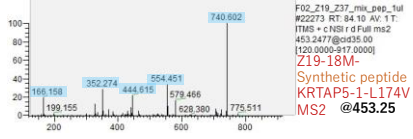


Z19 Clinical sample KRTAP5-1-L174V Scaffold result

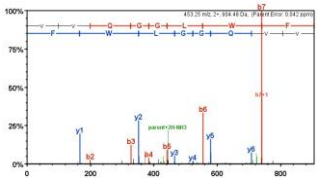
NL: 2.64e3



Z19Clinical sample KRTAP5-1-L174V MS2 @453.247



Z19-18M-Synthetic peptide KRTAP5-1-L174V MS2 @453.25

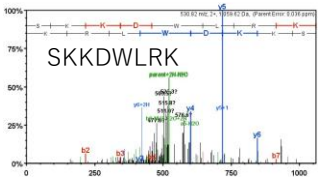


Z19-18M-Synthetic peptide KRTAP5-1-L174V Scaffold result

Peptide ID: Z34-2M | Relative retention time: 1.3

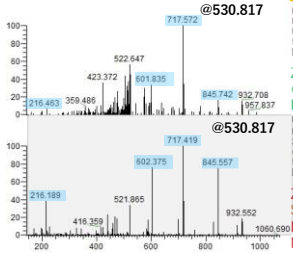
Z34-2M-FAM221A-Indel_167 Hp: 16.68

Mascot: 14.9



Z34 Clinical sample FAM221A-Indel_167 Scaffold result

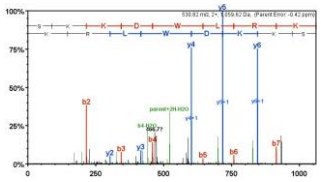
NL: 1.79e3



Z34Clinical sample のFAM221A-Indel_167 MS2

Consistent peak

Z34-2M-Synthetic peptide FAM221A-Indel_167 MS2

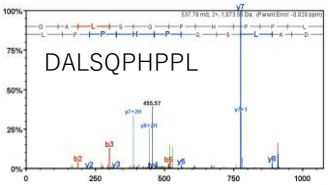


Z34-2M-Synthetic peptide FAM221A-Indel_167 Scaffold result

Peptide ID: Z35-13M | Relative retention time: 1.3

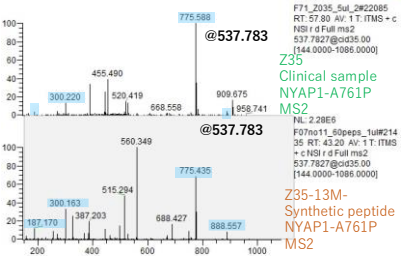
Z35-13M-NYAP1-A761P Hp: 21.58

Mascot: 14.5



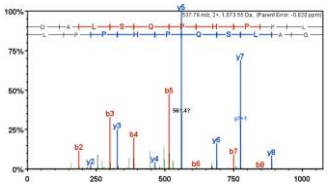
Z35 Clinical sample NYAP1-A761P Scaffold result

NL: 2.15e4



Z35 Clinical sample NYAP1-A761P MS2

Z35-13M-Synthetic peptide NYAP1-A761P MS2

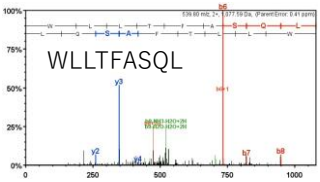


Z35-13M-Synthetic peptide NYAP1-A761P Scaffold result

Peptide ID: Z35-16M | Relative retention time: 0.9

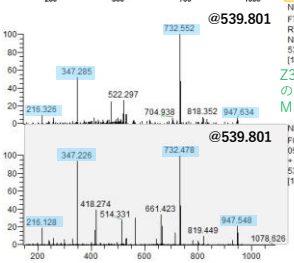
Z35-16M-SLC12A4-R712W Hp: 41.82

Mascot: 11.5



Z35 Clinical sample SLC12A4-R712W Scaffold result

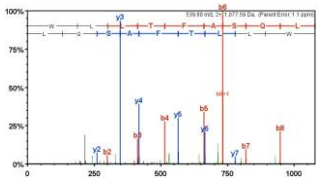
NL: 1.02e4



Z35Clinical sample のSLC12A4R712W MS2

Consistent peak

Z35-16M-Synthetic peptide SLC12A4-R712Wの MS2

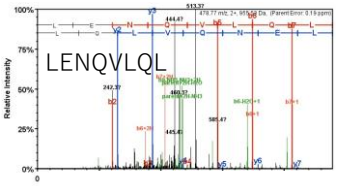


Z35-16M-Synthetic peptide SLC12A4-R712W Scaffold result

Peptide ID: Z37-2M | Relative retention time: 1.2

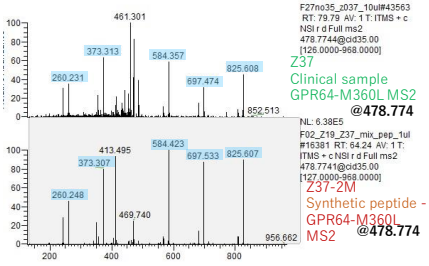
Z37-2M-GPR64-M360L Hp: 26.64

Mascot: 19.3



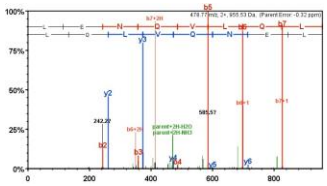
Z37 Clinical sample GPR64-M360Lの Scaffold result

NL: 2.58e3



Z37 Clinical sample GPR64-M360L MS2 @478.774

Z37-2M Synthetic peptide - GPR64-M360L MS2 @478.774



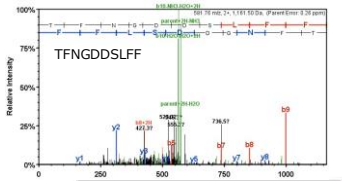
Z37-2M-Synthetic peptide GPR64-M360L Scaffold result

Peptide ID: Z37-14M | Relative retention time: 0.9

Z37-14M-CFTR-S422F

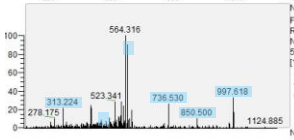
Hp: 33.96

Mascot: 11.7



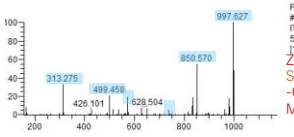
Z37 Clinical sample のCFTR-S422F Scaffold result

NL: 1.32e3

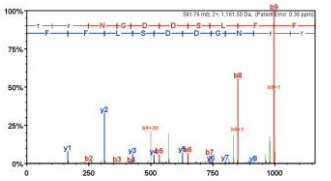


Z37Clinical sample CFTR-S422F MS2 @581.756

Consistent peak



Z37-14M Synthetic peptide -CFTR-S422F MS2 @581.757



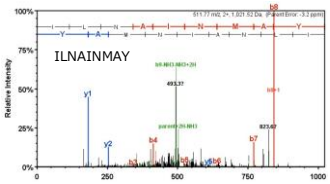
Z37-14M Synthetic peptide -CFTR-S422F Scaffold result

Peptide ID: Z37-23M | Relative retention time: 0.8

Z37-23M-P2RY2-T280I

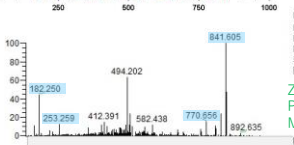
Hp: 29.60

Mascot: 11.2

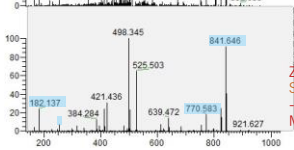


Z37Clinical sample P2RY2-T280I Scaffold result

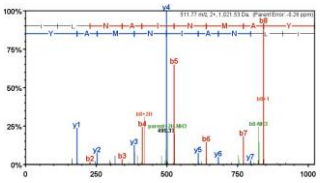
NL: 4.14e3



Z37Clinical sample P2RY2-T280I MS2 @511.769



Z37-23M Synthetic peptide -P2RY2-T280I MS2 @511.771



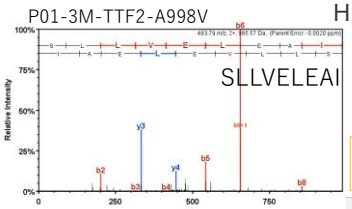
Z37-23M Synthetic peptide -P2RY2-T280I Scaffold result

Peptide ID: P01-3M | Relative retention time: 0.9

P01-3M-TTF2-A998V

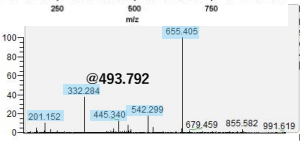
Hp: 37.11

Mascot: 28.6



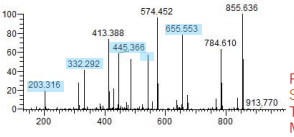
P01 Clinical sample TTF2-A998V Scaffold result

NL: 2.23e4

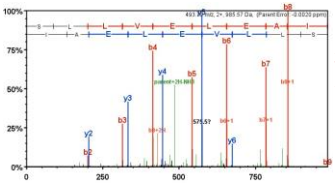


P01 Clinical sample TTF2-A998V MS2 @493.792

Consistent peak



P01-3M- Synthetic peptide TTF2-A998V MS2 @493.793



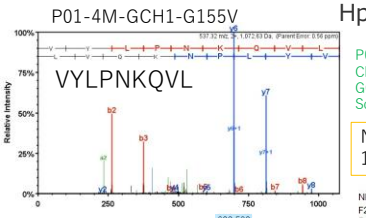
P01-3M- Synthetic peptide TTF2-A998V Scaffold result

Peptide ID: P01-4M | Relative retention time: 1.1

P01-4M-GCH1-G155V

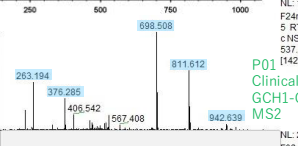
Hp: 28.19

Mascot: 25.6

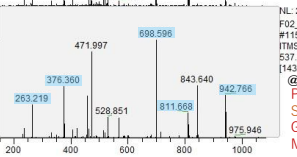


P01 Clinical sample GCH1-G155V Scaffold result

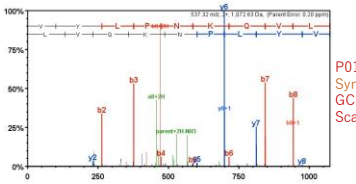
NL: 1.49e4



P01 Clinical sample GCH1-G155V MS2 @537.322



P01-4M Synthetic peptide GCH1-G155V MS2 @537.322

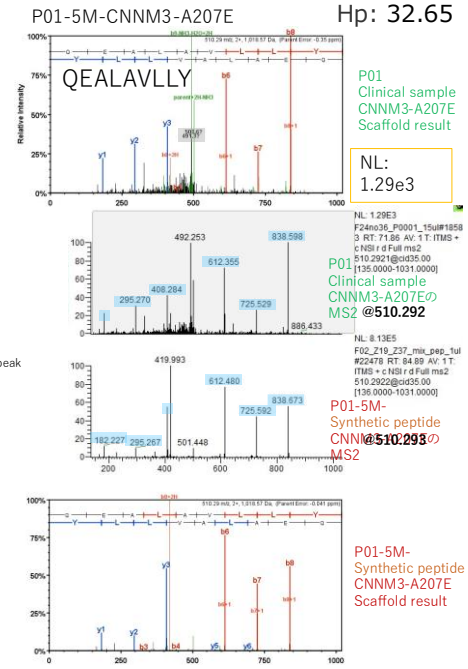


P01-4M Synthetic peptide GCH1-G155V Scaffold result

Peptide ID: P01-5M | Relative retention time: 0.8

P01-5M-CNNM3-A207E | Hp: 32.65

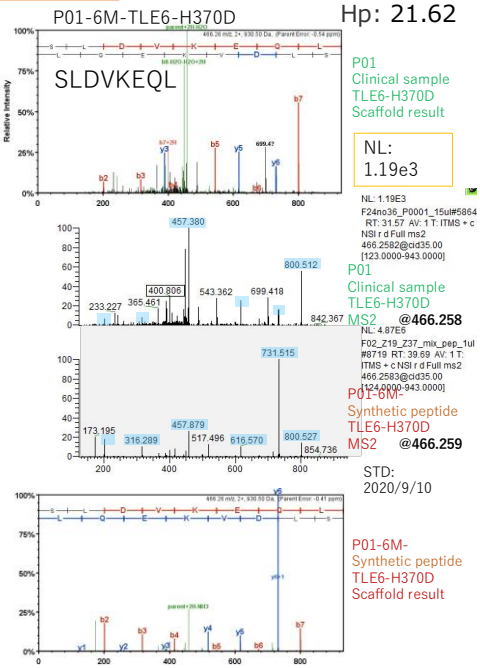
Mascot: 21.3



Peptide ID: P01-6M | Relative retention time: 0.8

P01-6M-TLE6-H370D | Hp: 21.62

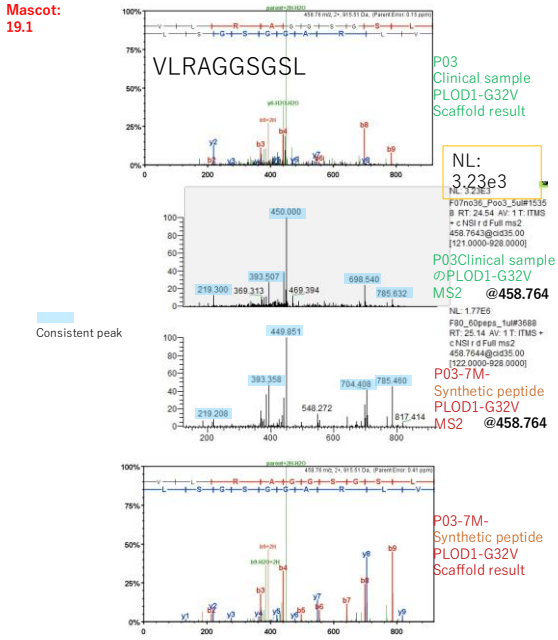
Mascot: 24.1



Peptide ID: P03-7M | Relative retention time: 1.0

P03-7M-PLOD1-G32V | Hp: 16.74

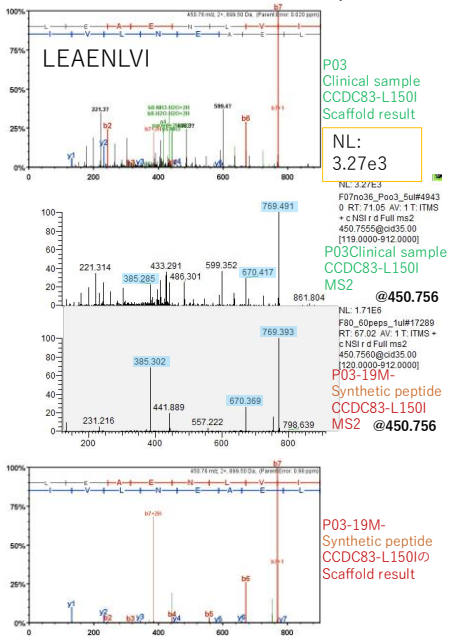
Mascot: 19.1



Peptide ID: P03-19M | Relative retention time: 1.1

P03-19M-CCDC83-L150I | Hp: 27.50

Mascot: 15.5



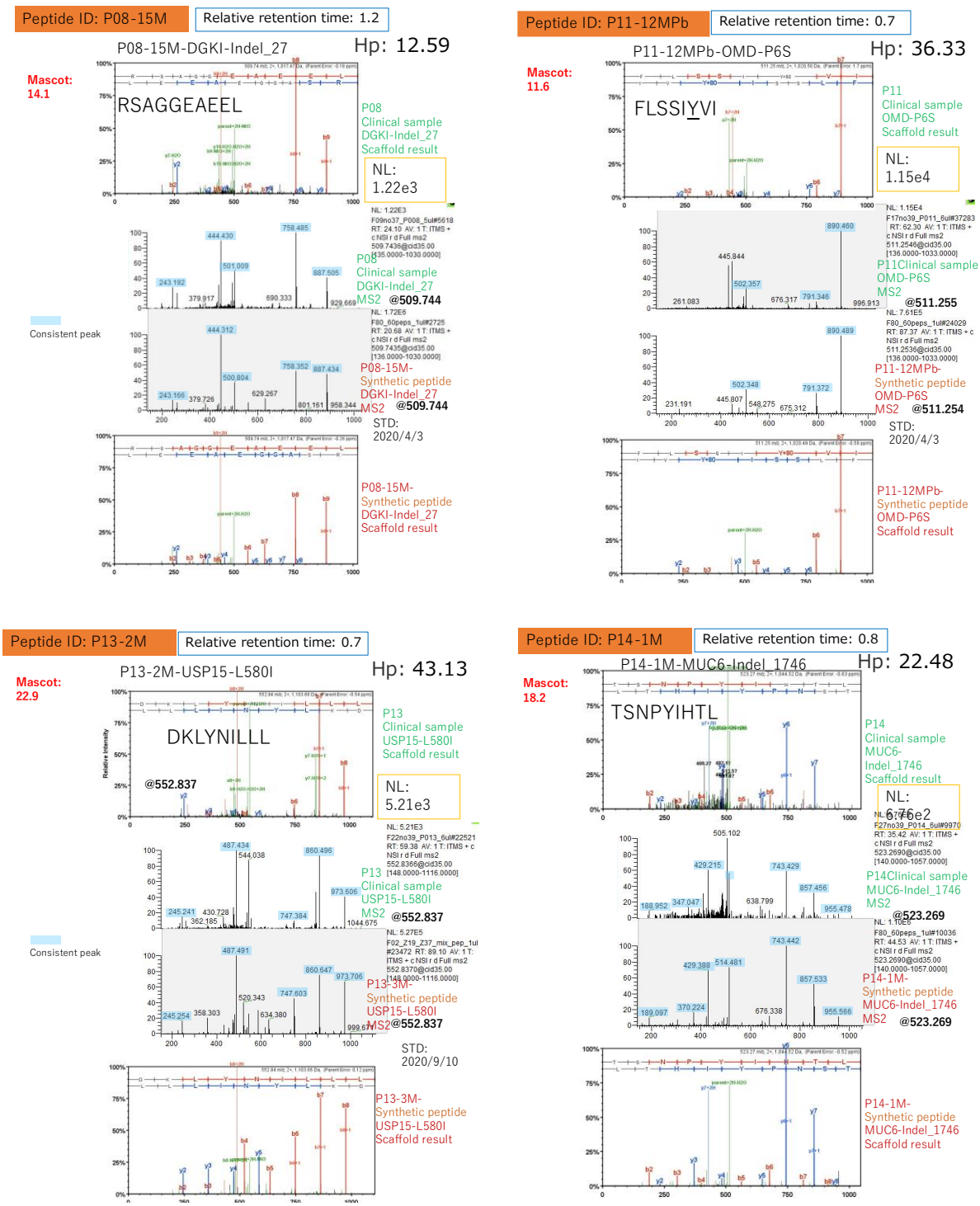


Figure S1. MS/MS spectra of HCS. HCS, reliable identifications with highly consistent MS/MS spectra.

a

x	y	R-squared	F-statistic	p-value
Hydrophobicity	SynRT_HSC	0.88	164.9	1.07E-11***
Gravy	SynRT_HSC	0.55	26.89	3.37E-05***
pI	SynRT_HSC	0.35	11.61	0.002523**
AliphaticIndex	SynRT_HSC	0.30	9.521	0.005403**
Entrp	SynRT_HSC	0.05	1.236	0.2783
InstabilityIndex	SynRT_HSC	0.05	1.175	0.2901
MinRank	SynRT_HSC	0.02	0.4739	0.4984

b

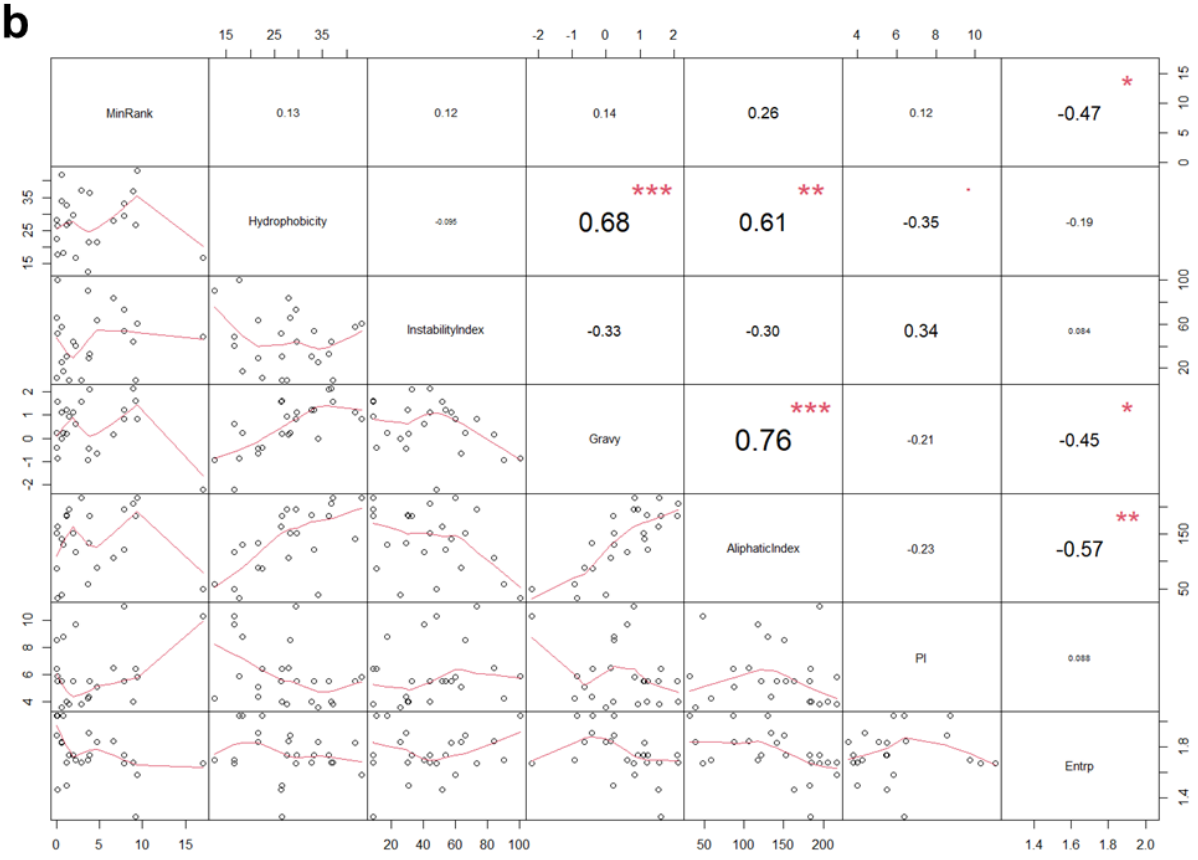


Figure S2. Selection of Hydrophobicity for predicting standard retention time: **a** outcome of univariate linear regression with the standard retention time (SynRT_HCS, y) for features predicted based on sequence information (x) across the 24 sequences exhibiting highly consistent MS/MS spectra (HCS). The features are ranked based on the R-squared value of the regression

line, and the linearity between Hydrophobicity and SynRT_HCS is optimal; **b** results of Spearman correlation analysis between features. Spearman coefficients plus the significance level were shown in the upper triangular, and the bivariate scatter plots with a fitted line were displayed in the lower part below the diagonal. Features exhibiting considerable linearity with SynRT_HCS (Gravy, pI and AliphaticIndex) display significant or strong correlations with Hydrophobicity. Considering the limited sample size of HCS and the risk of overfitting caused by multicollinearity in the multivariable regression model, the univariable linear regression model (as shown in **Figure 2**) using Hydrophobicity was selected to give a relatively robust performance. Significance level: *** P-value 0 - 0.001; ** P-value 0.001 - 0.01; * P-value 0.01 - 0.05; • P-value 0.05 - 0.10.

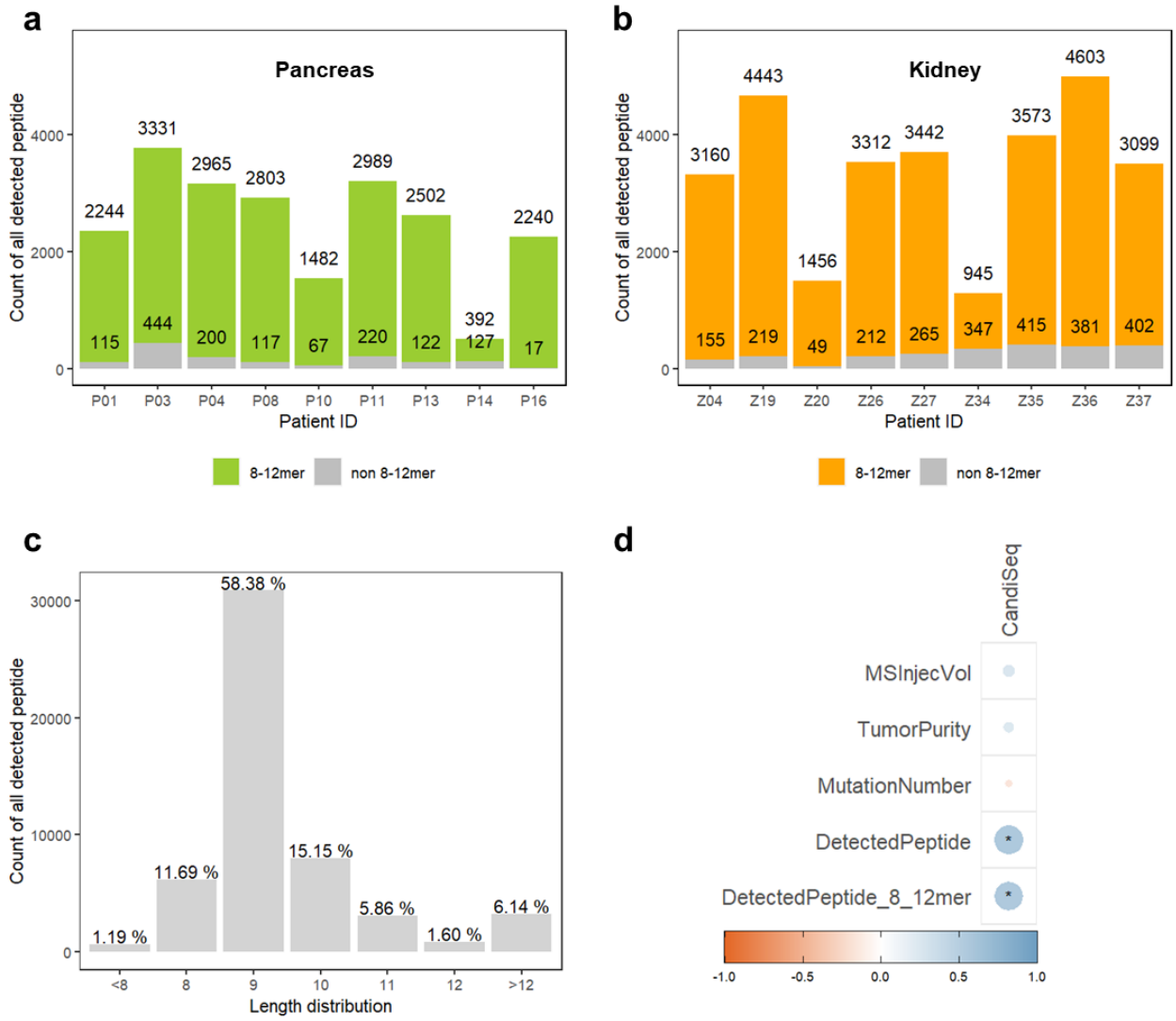


Figure S3. Count of overall detected peptides (Mascot ≥ 10) from patients with **a** pancreas and **b** kidney cancer; **c** the length distribution of all detected peptides; **d** Spearman's correlation coefficients between count of CandiSeqs (CandiSeq), tumor tissue input (MSInjecVol), tumor purity (TumorPurity), mutation number (MutationNumber), count of overall detected peptides (DetectPeptide), and count of 8-12mers in overall detected peptides (DetectedPeptide_8_12mer). Number labels in **a** and **b**, count number; number labels in **c**, percentage in total detected peptides; circle size in **d**, |Spearman's correlation coefficient|; circle color, red: negative, blue: positive value; * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001.