Supplementary Information

Efficient discovery of anticancer peptides via cost-aware ranking learning

Jianda Yue [a,b,c,d,#], Jiawei Xu [a,b,c,d,#], Zihui Chen [a,b,c,d], Tingting Li [a,b,c,d], Zhaoyang Tang [a,b,c,d], Xie Li [a,b,c,d], Hua Tan [a,b,c,d], Wangfei Xiang [a,b,c,d], Zhonghua Liu [a,b,c,d], Ying Wang [a,b,c,d,*]

[a]The National and Local Joint Engineering Laboratory of Animal Peptide Drug Development, College of Life Sciences, Hunan Normal University, Changsha 410081, China

[b]Hunan Xiangjiang Laboratory, Changsha 410205, China

[c]Peptide and small molecule drug R&D plateform, Furong Laboratory, Hunan Normal University, Changsha 410081, China

[c]Institute of Interdisciplinary Studies, Hunan Normal University, Changsha 410081, China

[#]The authors give equal contributions

*Corresponding author

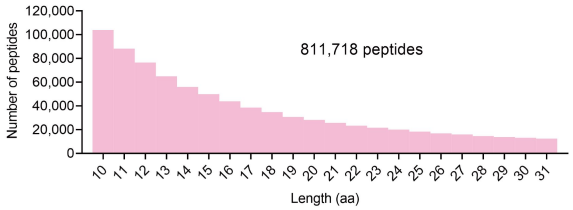
Supplementary Fig. 1 | Visualization of ranking performance on the external rPep dataset (Part I). The figure displays the prioritized lists generated by **(a)** ACPRank and **(b)-(p)**, 15 comparative mainstream ACP prediction models. Sequences are sorted from left to right based on the predicted probability scores output by each model. The number inside each box represents the peptide's ground-truth activity rank. Red backgrounds highlight the positions of the Top 3 elite targets. Blue-framed boxes indicate samples that were classified as non-ACPs (negative) by the specific binary decision threshold of each model.

Supplementary Fig. 2 | Visualization of ranking performance on the external rPep dataset (Part II). This figure continues the benchmark comparison, displaying the ranking lists generated by (**a**)-(**o**), 15 additional ACP prediction models. The visualization scheme is identical to **Supplementary Fig. 1**.

a

Tissue	Substrate		Protease		Peptide	
	Count	Pct. (%)	Count	Pct. (%)	Count	Pct. (%)
Wide expression	318	14.94	8	8.51	14,594,033	19.09
Brain	214	10.05	2	2.13	8,466,651	11.07
Liver	207	9.72	18	19.15	8,442,585	11.04
Intestine	224	10.52	11	11.7	6,725,916	8.8
Placenta	108	5.07	5	5.32	5,789,343	7.57
Lymphoid tissue	281	13.2	7	7.45	5,123,828	6.7
Salivary gland	109	5.12	8	8.51	5,033,740	6.58
Cervix	69	3.24	4	4.26	3,808,478	4.98
Testis	88	4.13	3	3.19	3,389,332	4.43
Retina	64	3.01	3	3.19	3,379,260	4.42
Adipose tissue	55	2.58	3	3.19	3,352,913	4.38
Skin	105	4.93	8	8.51	3,204,043	4.19
Kidney	53	2.49	3	3.19	2,628,419	3.44
Ovary	54	2.54	3	3.19	2,598,660	3.4
Esophagus	92	4.32	10	10.64	2,538,808	3.32
Stomach	104	4.88	4	4.26	2,299,908	3.01
Heart muscle	44	2.07	2	2.13	2,143,375	2.8
Lung	54	2.54	1	1.06	2,042,124	2.67
Pituitary gland	50	2.35	0	0	2,030,749	2.66
Epididymis	93	4.37	2	2.13	1,882,157	2.46
Choroid plexus	61	2.87	1	1.06	1,825,766	2.39
Thyroid gland	33	1.55	1	1.06	1,794,395	2.35
Bone marrow	87	4.09	7	7.45	1,781,258	2.33
Gallbladder	42	1.97	2	2.13	1,746,366	2.28
Pancreas	63	2.96	6	6.38	1,734,178	2.27
Urinary bladder	97	4.56	4	4.26	1,641,422	2.15
Endometrium	28	1.32	1	1.06	1,561,187	2.04
Adrenal gland	35	1.64	2	2.13	1,537,485	2.01
Fallopian tube	15	0.7	0	0	1,461,329	1.91
Vagina	39	1.83	4	4.26	1,281,019	1.68
Smooth muscle	14	0.66	1	1.06	1,156,607	1.51
Parathyroid gland	28	1.32	2	2.13	1,137,314	1.49
Skeletal muscle	36	1.69	0	0	919,530	1.2
Prostate	20	0.94	5	5.32	908,651	1.19
Seminal vesicle	18	0.85	0	0	747,181	0.98
Breast	33	1.55	1	1.06	742,547	0.97
Tongue	17	0.8	0	0	570,400	0.75
No annotation	155	7.28	1	1.06	2,861,083	3.74
Total	2,129		94		76,464,332	

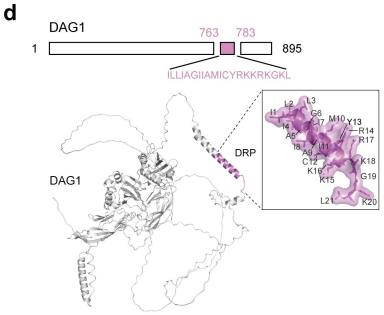
b



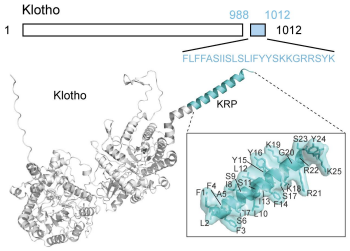
c

Sequence	Uniprot ID		Rank	Length	Charge	GRAVY
	Protease	Substrate				
QPPPPFTAPM	Q9BYF1	Q14118	7	10	-0.24	-0.57
TPINPMIPGL	P00742	Q8TC92	11	10	-0.6	0.53
SYDYSYWDGS	P00742	Q6UXS0	13	10	-2.54	-1.46
PSTNLQNIPM	P00742	Q9NR99	16	10	-0.04	-0.50
HNSAPMFMLDLYNAMAVEEG	P00742	P18075	19	20	-3.15	0.05
HGTLYPPGSV	P00797	Q6ZRI0	23	10	-0.15	-0.20
YMLYMQMISL	Q92876	P0DTE2	24	10	-0.24	1.09
CDSGGDCECFCTAVAAYAQACHDAGL	Q96IY4	Q9HC84	26	26	-4.2	0.36
PQDPSFQMPL	P0DJD9	Q08380	27	10	-1.04	-0.76
KDPGALMFPI	Q9GZN4	P45452	28	10	-0.24	0.38
QGPLNSWIPPF	P07477	A1E959	29	11	-0.24	-0.26
GYNQILIVPM	Q14CN2	O95428	30	10	-0.24	0.86
YIPYIILTIIILTIVGFIWLLKVNG	A8K7I4	Q8TD46	31	25	0.76	1.91
IFLHFSLLTCLSWMGLEGYNLYR	Q92824	Q9Y653	32	23	-0.16	0.70
QGPGIECTGFCAPGCTCPPGLFLHN	P09958	A2VEC9	33	25	-1.19	0.24
LIVGIVVGLLLAALVAGVVYWLYMKKSK	Q9UKR0	Q13740	34	28	2.76	1.68
PDFDPALGMM	Q9UNI1	Q16206	35	10	-2.04	0.16
ILLIAGIIAMICYRKKRKGKL	Q12884	Q14118	36	21	5.75	0.72
DGTFMPSIDVT	Q9UKR3	Q9H9K5	37	11	-2.24	0.20
PTPSGFEPTV	O15393	A6NGW2	41	10	-1.04	-0.39
PAGGGVGAGAAAVASAGVIYPYQPRAR	Q9UKR0	Q08397	43	27	1.96	0.31
ASTLSATPIDTSTPVTTSTEAT	P04070	Q685J3	45	22	-2.2	-0.05
YFDGQYWLWWV	Q9GZN4	Q9BT67	46	11	-1.24	-0.17
PVIPQQPMMP	Q14CN2	Q99217	47	10	-0.04	-0.09
FLFFASIISLSLIFYYSKKGRRSYK	Q9Y5K2	Q9UEF7	49	25	4.75	0.36

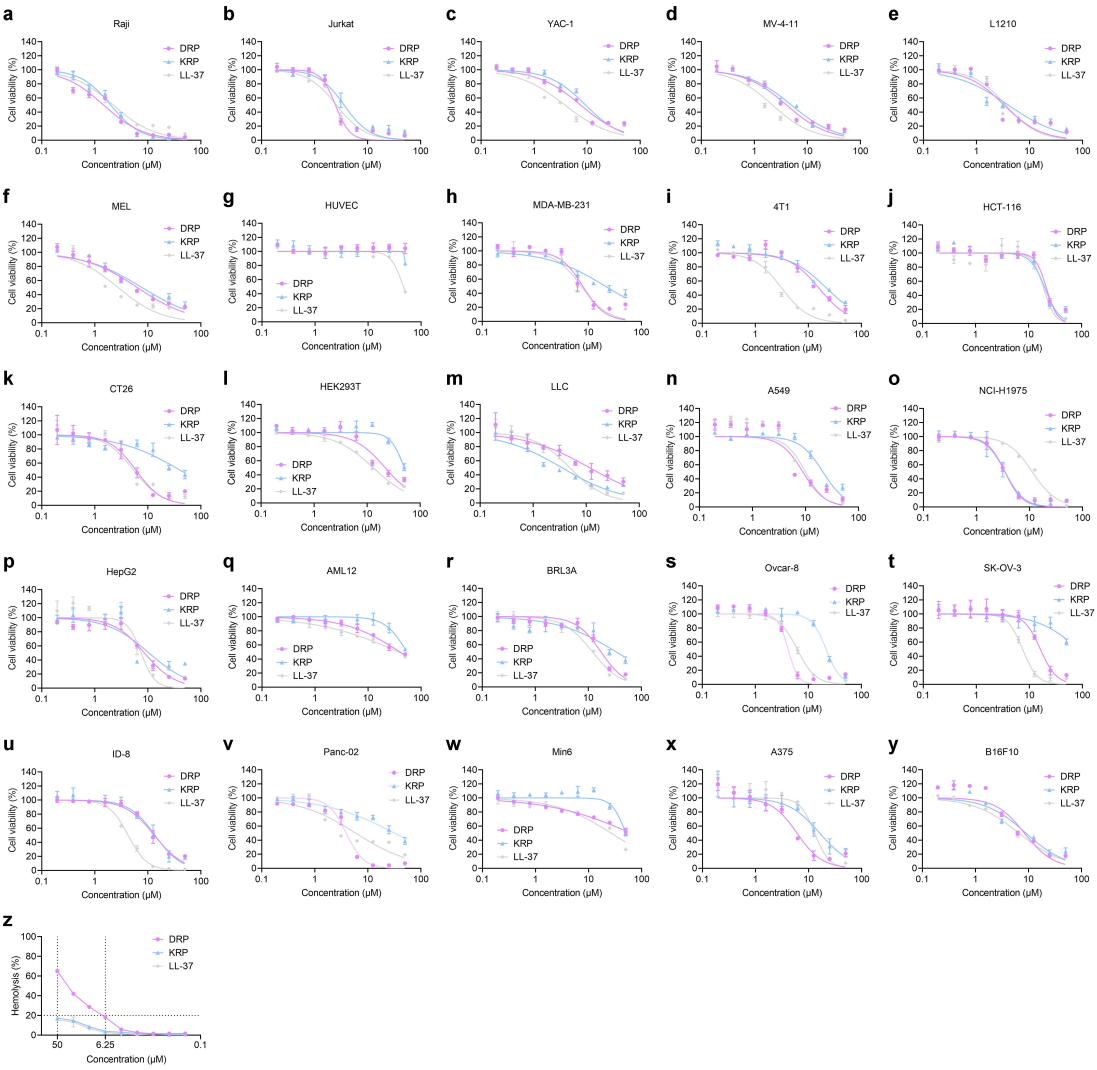
d



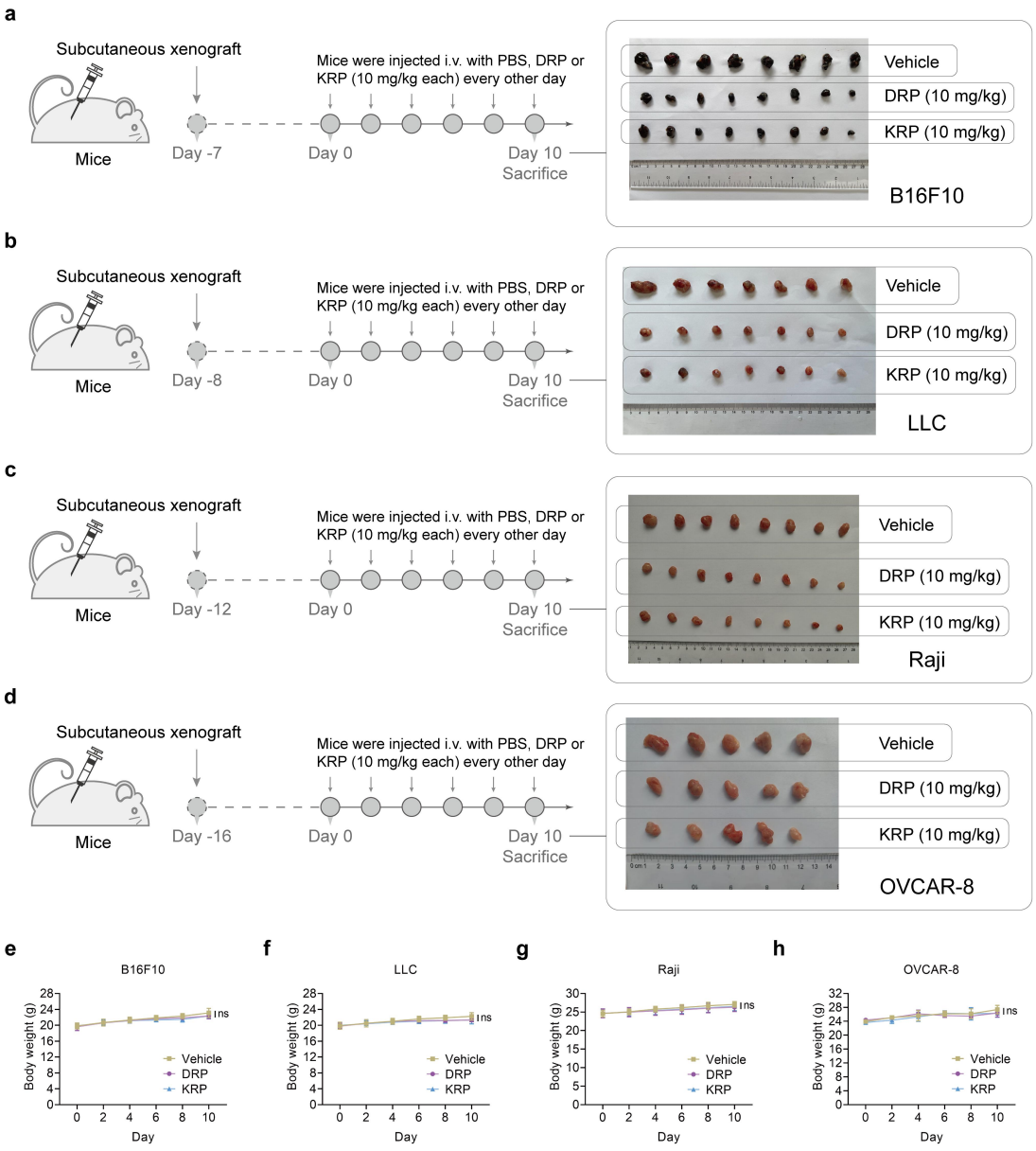
e



Supplementary Fig. 3 | Construction of the *in silico* peptide library and identification of DRP and KRP. (a) Tissue distribution statistics of the 2,129 secreted substrates, 94 secretory proteases, and the resulting initial peptide pool. The figure categorizes the count and percentage of proteins and generated peptides across various human tissues. **(b)** Length distribution histogram of the final 811,718 candidate peptides established after redundancy removal. **(c)** List of the top 25 candidate peptides retained after ACPRank prediction and similarity filtering (Needleman-Wunsch algorithm, similarity cutoff < 45%). The table details the UniProt IDs of the corresponding proteases and substrates, along with key physicochemical properties including rank, peptide length, net charge, and GRAVY scores. The two selected candidates, DRP (rank 36) and KRP (rank 49), are highlighted in bold. **(d, e)** Schematic representation of the sequence location and 3D structural mapping of DRP within the DAG1 **(d)** and KRP within the Klotho **(e)**. The zoomed-in panels illustrate the specific amino acid sequences and helical structures of the derived peptides.



Supplementary Fig. 4 | *In vitro* cytotoxicity and hemocompatibility profiles. (**a–y**) Dose–response curves illustrating the cell viability of 25 cell lines treated with varying concentrations of DRP, KRP, or LL-37. **(z)** Hemolysis assay of murine erythrocytes incubated with the indicated peptides.



Supplementary Fig. 5 | *In vivo* therapeutic efficacy and body weight evaluation of DRP and KRP. (a–d) Schematic illustration of the experimental design (left) and representative photographs of excised tumors at the study endpoint (right) for B16F10 **(a)**, LLC **(b)**, Raji **(c)**, and OVCAR-8 **(d)** xenograft models. Following tumor establishment, mice received intravenous injections of vehicle (PBS), DRP, or KRP (10 mg/kg) every other day as indicated in the timeline. **(e–h)** Body weight monitoring curves of B16F10 **(e)**, LLC **(f)**, Raji **(g)**, and OVCAR-8 **(h)** tumor-bearing mice throughout the treatment course. "ns" indicates not significant.

Supplementary Table 1 | List of unseen query sequences used in the anchor insertion stress test.

Peptide	Name	Sequence
tPep1	Cycloviolacin O2	GIPCGESCVWIPCISSAIGCSCKSKVCYRN
tPep2	Cliotides T12, CT12	GIPCGESCVYIPCTVTALLGCSCKDKVCYKN
tPep3	Vitri F	GTLPCGESCVWIPCISSVVGCACKSKVCYKD
tPep4	Mram 8	GIPCGESCVFIPCLTSAIGCSCKSKVCYRN
tPep5	Cliotides T10, CT10	GVPCAESCVWIPCTVTALLGCSCKDKVCYLN
tPep6	FR-15	FRRFFKWFRRFFKFF
tPep7	B9	GNPCGESCVYLPCITTVVGCSCQNSVCYHN
tPep8	Hymenochirin-1Pa	LKLSPKTKDTLKKVLKGAIKGAIAIASMA
tPep9	Viphi A	GSIPCGESCVFIPCISSVIGCACKSKVCYKN
tPep10	Hymenochirin-1B	IKLSPETKDNLKKVLKGAIKGAIAVAKMV
tPep11	[K8R]cGm	GCRRLCYRQRCVTYCRGR
tPep12	Viphi F	GSIPCGESCVFIPCISAIIGCSCSSKVCYKN
tPep13	FR11P	FRRFFKWFRRPFKFF
tPep14	CA(1-& HECATE(11/23)	KWKLFKKALKKLKKALKKAL
tPep15	A12L	KWKSFLKTFKSLKKTVLHTALKAISS
tPep16	Maximin 4	GIGGVLLSAGKAALKGLAKVLAEKYAN
tPep17	Brevinin-1H	FALGAVTKVLPKLFCLITRKC
tPep18	Phakellistatin 7	PPIFALPPYI
tPep19	Hecate Ac	FALALKALKKALKKLKKALKKAL
tPep20	Kalata B1 [N15Y]	GLPVCGETCVGGTCYTPGCTCSWPVCTRN