In vivo phage display: identification of organ-specific peptides using deep sequencing and differential profiling across tissues

*Karlis Pleiko[1,2], *Kristina Põšnograjeva[1], Maarja Haugas[1], Kaarel Kurm[1], Una Riekstina[2], Tambet Teesalu[1,3,4]

* These authors contributed equally to this work

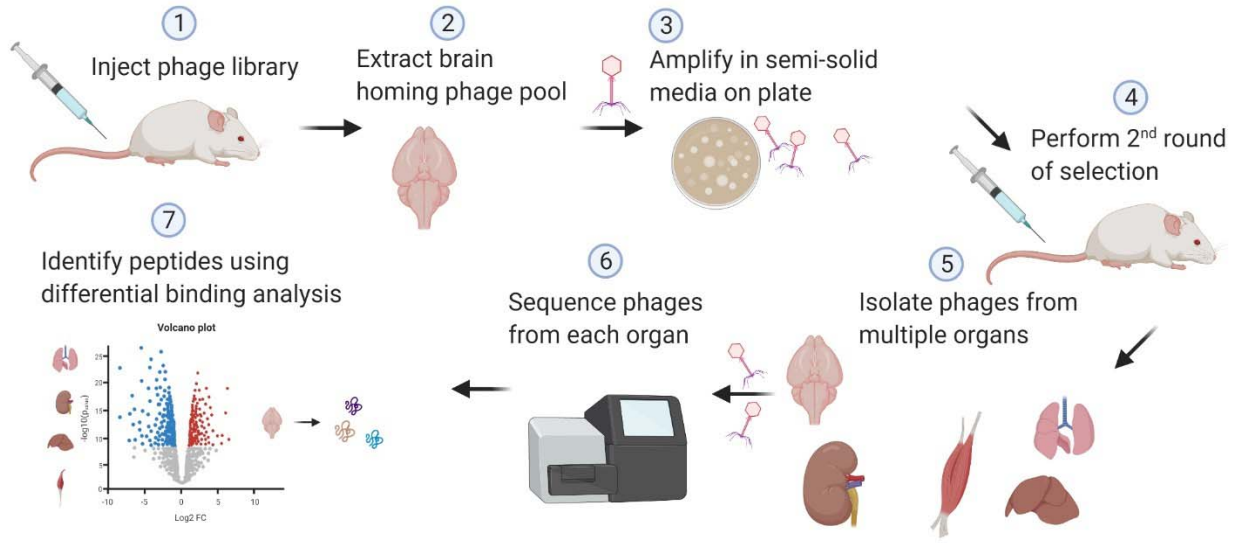
[1] Laboratory of Cancer Biology, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia

[2] Faculty of Medicine, University of Latvia, Riga, Latvia

[3] Cancer Research Center, Sanford Burnham Prebys Medical Discovery Institute, La Jolla, California, USA

[4] Center for Nanomedicine and Department of Cell, Molecular and Developmental Biology, University of California, Santa Barbara, Santa Barbara, California, USA

Graphic abstract



In vivo phage display using differential binding approach

ABSTRACT

In vivo phage display is widely used for identification of organ- or disease-specific homing peptides. However, the current *in vivo* phage biopanning approaches fail to assess biodistribution of specific peptide phages across tissues during the screen, thus necessitating laborious and time-consuming post-screening validation studies on individual peptide phages. Here, we adopted bioinformatics tools used for RNA sequencing for analysis of high throughput sequencing (HTS) data to estimate the representation of individual peptides during biopanning *in vivo*. The data from *in vivo* phage screen were analyzed using differential binding – relative representation of each peptide in the target organ vs. in a panel of control organs. Application of this approach in a model study using low-diversity peptide T7 phage library with spiked-in brain homing phage, demonstrated brain-specific differential binding of brain homing phage and resulted in identification of novel lung and brain specific homing peptides. Our study provides a broadly applicable approach to streamline in vivo peptide phage biopanning and to increase its reproducibility and success rate.

INTRODUCTION

The ability to present foreign peptides as fusions to the nucleocapsid proteins on the surface of bacteriophage was first described by Smith in 1985 (1). Since then, application of the engineered phages for *de novo* peptide ligand discovery has become invaluable and widely applied tool (2). Pasqualini and Ruoslahti introduced *in vivo* phage display in 1996 (3) and adopted the method for identification of peptides that target tumor vasculature in 1998 (4). During *in vivo* biopanning, after short circulation (typically 10 min to 1 h) free phages and phages expressing weakly binding peptide sequences are removed by perfusion. Target tissues, e.g. xenograft tumors, or specific organs, are then collected and tissue-bound phages are rescued by amplification for use in subsequent round of selection (5).

A number of tumor-homing peptides have been identified using *in vivo* phage display. The power of in vivo display is particularly well illustrated by identification of a novel class of tumor homing peptides, tumor penetrating peptides, that use a complex multistep mechanism, very hard to envision to be designed rationally, for tumor-specific homing and vascular exit. The prototypic tumor penetrating peptide, iRGD (amino acid sequence: CRGDKGPDC) contains a RGD motif that binds the integrins and a cryptic R/KXXR/K C-end Rule (CendR) motif that is activated by proteolytic cleavage (6). C-terminally exposed CendR motif binds to neuropilin-1 (NRP-1) to trigger endocytic transport and tumor penetration pathway named CendR pathway (7). The use of CendR pathway can improve transport conjugated and coadministered therapeutic compounds into tumors *in vivo* (8). iRGD peptide is clinically developed by CEND Therapeutics Inc. (La Jolla, CA, USA) in phase I clinical trial on pancreatic cancer patients as an add-on treatment in combination with Abraxane and gemcitabine (clinical trial identifier: NCT03517176). Other vascular-homing peptides Cilengitide (sequence: RGDfV) (9), SFITGv6 (sequence: GRCTFRGDLMQLCYPD) (10), CNGRC (sequence: CNGRC) (11), tumor extracellular matrix (ECM) homing peptides DAG (sequence: CDAGRKQKC) (12), ZD2 (sequence: CTVRTSADC) (13), PL1 (sequence: PPRRGLIKLKTS) (14) and PL3 (sequence: AGRGRLVR) (15), IP3 (sequence :CKRDLSRRC) (16), CAR (sequence: CARSKNKDC) (17), and M2 tumor associated macrophage targeting UNO peptide (18) have been selected for possible therapeutic use and reviewed recently. Moreover, healthy organs can be specifically targeted with homing peptides. Examples are brain homing peptide CAGALCY (19, 20) and

prototypic C-end Rule (CendR) peptide RPARPAR that upon intravenous administration accumulates in lungs and heart (7).

High-throughput sequencing (HTS) of peptide-encoding region of the phage genome can be used to gain insight into the evolution of the peptide landscape during biopanning. HTS has proven particularly useful for binding peptide identification during “low noise” selections on purified individual target proteins (21). However, HTS-based identification of *in vivo* homing peptides remains challenging. One reason is that the peptides that accumulate in intended target tissue can also home to other sites, including healthy organs. Such off-target binding will only be noted when biodistribution of individual phages is assessed. Another complicating factor, especially when using *in vivo* phage display with lysogenic filamentous phages that rely on bacterial cellular machinery for export, is the enrichment of parasitic peptide sequences that increase phage fitness and selection against phages displaying peptides that reduce phage amplification. During M13 phage amplification representation of different peptide phages in a pool can be skewed 10-fold to 100-fold (22). Two approaches can be used to reduce the amplification bias – either avoiding the amplification altogether or performing the amplification in a way that each phage has the same amplification ratio (21, 23). Bioinformatics tools have been developed to prospectively identify parasitic peptide sequences, such as SAROTUP (24) and PhD7Faster (25).

In recent years, a number of studies have been published on HTS-based phage display. Hurwitz et al. developed motif search algorithm that breaks 12-mer peptide sequences into smaller fragments and aligns them with other sequenced fragments using MAFFT software (26). Another motif search algorithm for mapping consensus sequences within the phage display-derived HTS datasets was developed by Rebollo et al. (27). Biopanning Data Bank (BDB) is a large scale effort to combine together data sets acquired from phage display experiments with more than 95 next generation phage display data sets and biopanning data from 1540 published articles (28). PHage Analysis for Selective Targeted PEPtides (PHASTpep) has introduced analysis of both positive and negative selection and sequencing both resulting phage pools. HTS data can be further sorted based on their abundance in each pool (29).

Here we present a statistical analysis-based approach for HTS data mining to determine differential binding of peptide phages across multiple organs during *in vivo* biopanning. Our approach eliminates parasitic peptide sequences in the selected peptide phage pool and

minimizes false leads thereby reducing the number of individual peptide phages that need to be used for *in vivo* biodistribution studies.

MATERIALS AND METHODS

In vivo phage display

For *in vivo* biopanning, we used a model low-diversity (10^5) CX_7C T7 phage peptide library with spiked brain-homing peptide (CAGALCY (19, 20)) as a positive control. The library was purified by PEG8000 precipitation and filtered through 0.2 µm syringe filter. For first biopanning round, 100 µL of 5×10^9 CX_7C T7 library with spiked-in 5×10^4 CAGALCY in Phosphate Buffered Saline (PBS) was injected intravenously (i.v.) in Balb/C mice. After 30 min, the animals were anesthetized and perfused with 10-15 ml of PBS containing 1% BSA (PBS-BSA). The target and control organs (brain, lung, liver, kidney, skeletal muscle) were collected, transferred in 1 mL of LB-1% NP-40, and homogenized on ice using handheld homogenizer with replaceable plastic tips. Phages in tissue lysates were rescued and amplified in *E. coli* strain BLT5403 in semi-solid media, purified by PEG8000 precipitation and filtered through 0.2 µm filter. For 2nd round of selection, 3 brain samples or 3 lung samples from 1st round were mixed and 5×10^9 plaque forming units (pfu) in 100 µL were used for injection (n=3 both for lung selection and brain selection) using the same procedure as in 1st cycle. Tissues were dissected and processed identically to 1st cycle. Final libraries were amplified, purified by PEG8000 precipitation and used for sequencing.

The animal procedures were approved by the Committee of Animal Experimentation, Estonian Ministry of Agriculture in accordance with Estonian legislation and Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes (permits #48 and #159).

Determination of phage displayed peptide sequences

Peptide-encoding region of bacteriophage genome was amplified by PCR using Phusion Green Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific, F537L) (reaction volume: 25 µL, list of primers added as Supplementary Table S1). Cycling conditions: denaturation at 98 °C for 30 sec., followed by 25 amplification cycles (10 sec at 98 °C, 21 sec at 72 °C), and final elongation (72 °C for 5 min). PCR products were purified using AMPure XP Bead Based Next-Generation Sequencing Cleanup system (Beckmann Coulter, A63881) using 1,8 µL of beads per 1 µL of PCR products. Purified PCR products were quantified using Agilent Bioanalyzer 2100 Instrument using the High sensitivity DNA Kit (Agilent, 5067-4626). Ion Torrent Emulsion PCR

and enrichment steps were performed using Ion PGM HiQ View OT2 kit (Life technologies, A29900). HTS was performed using Ion Torrent™ Personal Genome Machine™ (ION-PGM) using Ion PGM HiQ View sequencing kit (Life technologies, A30044) and Ion 316v2 chips (Life Technologies, 448149). The FASTQ sequence files were converted to text files and translated using in-house developed python scripts.

Peptide data mining

edgeR (30) was used for normalization of peptide count (Input data available as Supplementary Tables S2-S4) in the samples and for statistical analyses. Peptides with <5 reads present in <2 replicate samples were not used for further analysis. Three biological replicates of peptides present in brain or lung samples versus input library were compared for peptide representation analysis. Three biological replicates of peptides present in brain or lung samples versus control organs (liver, kidney, muscle, brain or lung) were compared for differential binding analysis. False discovery rate (FDR) cutoff of 0.05 and logFC cutoff of 2 were used as thresholds for statistical significance. UpSetR (31) package was used to create visualizations of the number of peptides in multiple-organ comparisons. Packages from tidyverse were used to create other sequencing analysis-related graphics (32).

Peptide phage cloning

Nucleotide sequences coding for BH1 (CLNSNKTNC), BH2 (CWRENKAKC) and LH1 (CGPNRRGDC) peptides were cloned into phage genomic DNA to be expressed on the phage surface as C-terminal fusions to capsid protein. The cloning was performed using complementary oligonucleotides (AATTCTTGCGGCCCGAACCGCCGCGGCGATTGCTA and AGCTTAGCAATCGCCGCGGCGGTTCGGGCCGCAAG for LH1, AATTCTTGCCTGAACA GCAACAAAACCAACTGCTA and AGCTTAGCAGTTGGTTTTGTTGCTGTTCAGGCAAG for BH1, AATTCTTGCTGGCGCGAAAACAAAGCGAAATGCTA and AGCTTAGCATTTC GCTTTGTTTTCGCGCCAGCAAG for BH2). Complementary oligonucleotides were diluted in mQ water at 100 nM, heated at 95°C for 5 min and slowly cooled to room temperature for annealing. The cloning was performed according to the manufacturer's protocol using the T7Select® 415-1 Cloning Kit (Millipore, 70015-3) with 1 µL of annealed oligonucleotides.

Intra-animal validation of candidate phages

Phages expressing a BH1, BH2, LH1, CAGALCY or G7 (GGGGGGG) peptides were amplified in *E. coli* strain BLT5403, purified by PEG8000 precipitation, followed by cesium chloride gradient centrifugation, and dialyzed against PBS. The phages were mixed in equimolar ratio and 1×10^{10} pfu of phage mix was injected i.v. into Balb/C mice (n=3). After 30 min, the animals were anesthetized and perfused with PBS-BSA. The target and control organs (brain, lung, liver, kidney, skeletal muscle) were collected, the tissues were homogenized and phages in tissue lysates were amplified and purified by PEG8000 precipitation. The distribution of phages between target and control organs was determined by HTS.

Inter-animal validation of candidate phages

5×10^{9} pfu of phage expressing BH1, BH2, LH1, CAGALCY or G7 peptide was injected i.v. into Balb/C mice (n=3). After 60 min, the animals were anesthetized and perfused with PBS-BSA. The target and control organs (brain, lung, liver, kidney, skeletal muscle) were collected, homogenized, and the amount of phage in tissue lysates was determined by titering using *E.coli* strain BLT5615.

Database search for target-unrelated peptides

All sequences that were identified using differential binding approach were searched in SAROTUP (24) to identify target-unrelated peptides (TUPScan), peptides containing sequences known to cause phages to grow faster (PhDFaster 2.0) and already known sequences in BDB (MimoScan).

RESULTS

Identification of candidate brain and lung homing peptides using *in vivo* peptide phage display with low complexity library

Low-diversity naïve CX_7C T7 phage peptide library with spiked-in brain-homing peptide CAGALCY was used as the starting library for *in vivo* phage display screens. For round 1 of biopanning, the starting library was injected into mice (n=3) and after 30 min of circulation, the mice were perfused to remove free and weakly bound background phages. Target (lung, brain) and control organs (liver, kidney, skeletal muscle) were collected, phages from these organs were amplified and the samples were subjected to Ion-Torrent HTS. Starting library was sequenced in 3 technical replicates. Phage pools recovered from lung and brain were used for two separate 2nd biopanning rounds (n=3 for each). After round 2, all control and target organs were collected, and phage pools were sequenced (Fig. 1). Sequencing was also performed for input libraries.

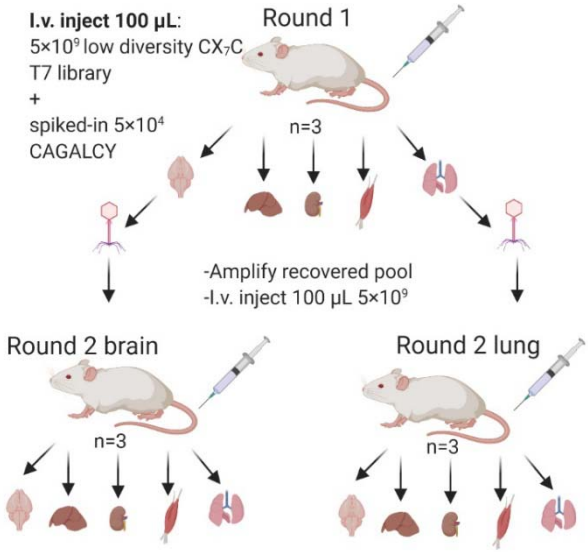


Fig. 1. Flow of in vivo phage biopanning. In round 1, 100 µL of 5×10^9 pfu collapsed CX_7C T7 phage library and 5×10^4 pfu of CAGALCY coding phage was injected in tail vein of mice (n=3). Target (lung and brain) and control organs (kidney, liver, muscle) were collected and phages were rescued by amplification in E. coli. For round 2, 5×10^9pfu of amplified libraries were injected in mice for further enrichment of lung (n=3) or brain (n=3) selective phages. All phage pools were subjected to HTS of the variable peptide encoding region of the phage genome.

The sequences of peptides expressed on the surface of phages recovered from target and control tissues were derived from the DNA sequencing data. The representation of reliable peptides, defined as peptides present in all replicates, was determined for three technical replicates of round 1 input library and for three biological replicates of phages amplified from organs (Fig. 2). At the sequencing depth used, reliable peptides in round 1 input library made up ~30% of all peptides. Phage pools recovered from brain displayed smaller fraction of reliable peptides (~10%). Percentage of reliable peptides in brain was the lowest, while in other organs percentage of reliable peptides from round 1 and round 2 brain samples were comparable to the one in technical replicates. Round 2 lung samples contained a smaller fraction of reliable peptides.

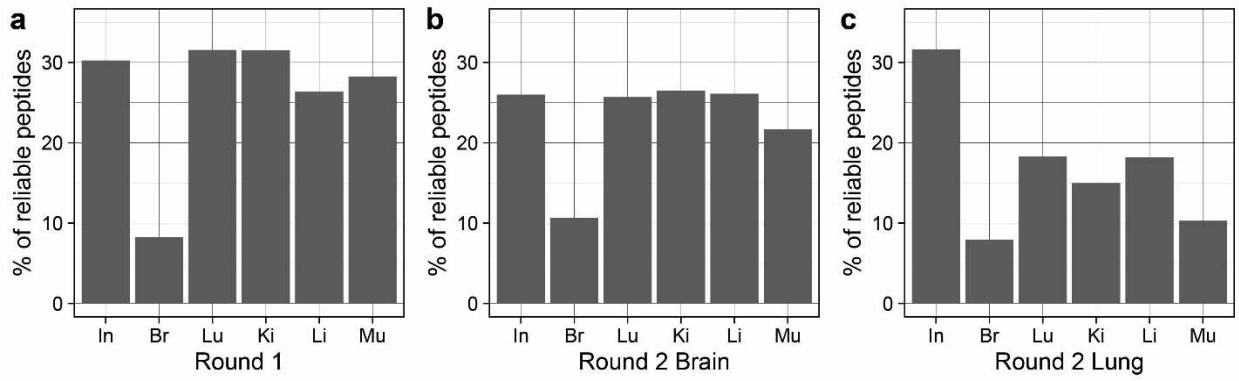


Fig. 2. Percentage of reliable peptides during in vivo selection rounds 1 and 2.

The graphs represent percentage of reliable peptides (defined as being present in n=3 replicates for each sample) in mice dosed with naïve ***(a)****, round 1 brain-selected* ***(b)****, or round 1 lung-selected libraries* ***(c)*** *. Input libraries represent technical variability between technical replicates (n=3) whereas organ-rescued libraries represent biological variability (n=3).*

Differential binding approach identifies organ-selective peptides

Organ selectivity of peptides was determined using differential binding approach, with statistical comparisons performed for each control organ (lung, kidney, liver, and muscle in the case the brain is the target organ) and for the target organ (brain) by using logFC>2 and FDR <0.05 as defined cutoffs. Each peptide that appeared brain-selective based on these criteria, was analysed further. Peptides from each comparison (e.g., lung-brain, kidney-brain, liver-brain and muscle-brain) were then cross-checked with all identified peptides in other comparisons. Target organ

selective peptides were defined as being present in all comparisons. Peptides that were selective only when compared to one or two control organs, were discarded (Fig. 3).

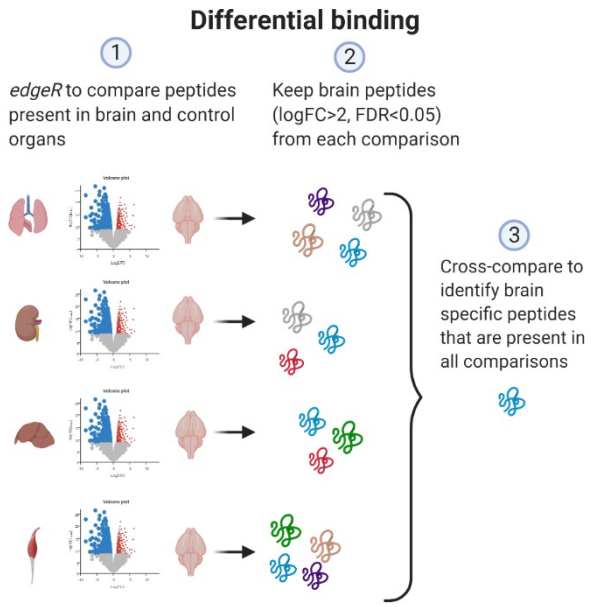


Fig. 3. Differential binding analysis.

HTS data from target organ (brain or lung) were compared to all control organs individually. Using edgeR, logFC > 2 and FDR <0.05 were applied as cutoff to identify peptides that are statistically signifficantly over-represented in brain. Peptides from each comparison were cross-compared to identify hits that appear in all comparisons as brain selective.

Landscapes of peptides displayed on phages recovered from brain samples in round 1 of biopanning were compared to those in the kidney, liver, muscle and lung. This analysis yielded a single brain-selective peptide when compared against the muscle sample. In round 2 of selection, 17 peptides were selective for the brain compared to the muscle, 9 compared to the liver, 7 compared to the kidney and 1 peptide (CAGALCY, known brain homing peptide spiked in the starting library) compared to the lung. CAGALCY was brain selective when compared to all control organs and 6 peptides appeared brain-selective when compared against kidney, liver and muscle (Fig. 4a). Based on logFC difference against control organs, BH1 (CLNSNKTNC) and BH2 (CWRENKAKC) were selected for individual testing as the most promising candidate brain homing peptides (Supplementary Table S5).

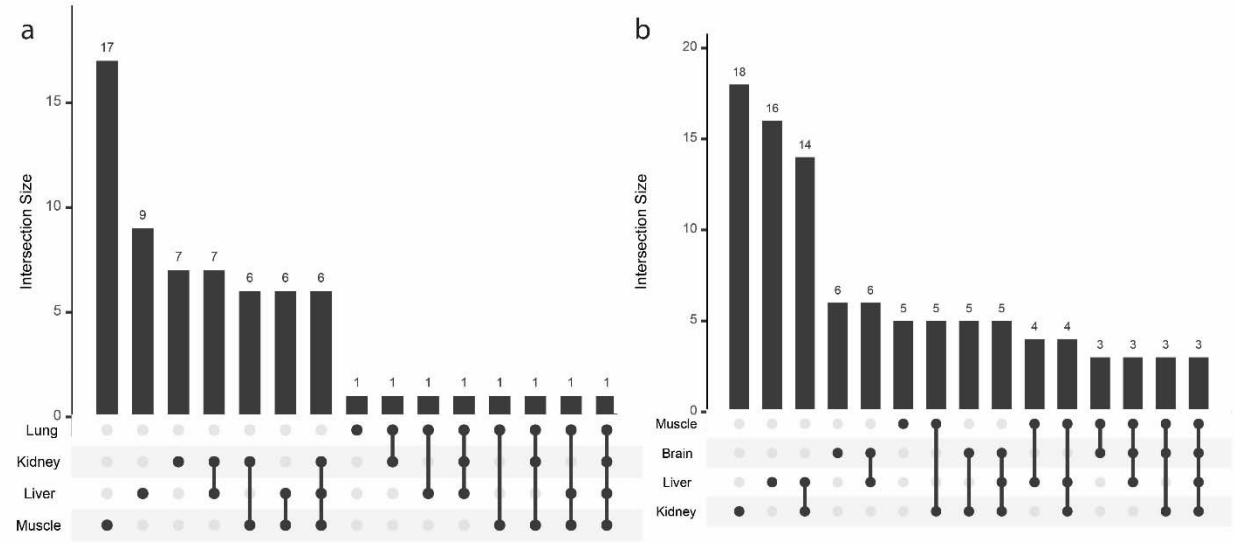


Fig. 4. Multiorgan comparison of selected phage pools to identify organ specific homing peptides.

Statistically significant (FDR < 0.05) peptides homing to the brain (a) and the lungs (b) compared against control organs (logFC > 2, FDR < 0.05) and number of peptides that overlap between different comparisons. One peptide (CAGALCY) is fully brain selective and 3 peptides lung selective when compared to other organs (Supplementary Tables S5 and S6).

Differential binding analysis for lung-selective peptides in mice dosed with naïve library yielded no lung selective peptides when compared to any other control organ. In contrast, round 2 yielded 18 lung selective peptides when the dataset was compared against the kidney, 16 peptides when compared against the liver, 6 when compared against the brain and 5 when compared against the muscle. Three peptides were present in all comparisons as lung-selective peptides (Fig. 4b). Notably, among 20 peptides that were lung selective compared to at least one organ, 8 contained the CX_5RGDC motif and 7 ended with C-terminal arginine residue (Supplementary Table S6). LH1 (CGPNRRGDC) was selected for further testing.

Enrichment analysis is the most commonly used approach to identify peptides in phage display. We next analyzed enrichment of reliable peptides in target organs in round 1 vs. round 2. Theoretically, the peptides that are most enriched, are most likely to reach the target and possess desired binding properties. We used gating logFC >2 and FDR < 0.05 between round 1 input

library and phages recovered from the target organs in rounds 1 and 2. As a threshold, we decided to apply the criteria used for differential binding analysis. After first round of selection, this approach resulted in identification of 4 peptides that are enriched in the lung and no brain-enriched peptides. After round 2, 16 peptides were found enriched in the brain (Supplementary Table S7) and 37 peptides in the lung (Supplementary Table S8).

One of 4 peptides enriched in the lungs after round 1 was an R-ending peptide, KTGARKR; the rest of sequences contained KLAALE phage vector-derived sequence. After round 2, among the 16 peptides enriched in the brain, CAGALCY was the most enriched sequence, followed by CVRLNKVRC (EN-2) and CTVTNKVRC (EN-3). EN-2 sequence in differential binding dataset shows selectivity only against the muscle and EN-3 sequence is selective for the brain when compared against the liver, kidney and muscle. Both BH1 and BH2 sequences that were identified using differential binding approach were present on the list as number 8 and number 4 accordingly. Of 37 peptides identified through enrichment analysis in lung, LH1 was 16th most enriched sequence.

Selected peptide phages show expected *in vivo* tropism in playoff analysis

In an earlier study, we developed a playoff analysis for the validation of candidate homing peptides (5). The method is based on medium-throughput internally controlled auditioning of mixtures of selected peptide phages. An equimolar mix of CAGALCY, BH1, BH2, LH1 and G7 control phages was i.v. injected into mice (n=3). Control organs (kidney, liver, muscle) and target organs (brain, lung) were collected and phage pools in each organ were sequenced using HTS. The input mixture was also sequenced. After normalization against representation in the input, fold change over G7 phage was calculated for each peptide phage in each organ (Fig. 5).

The phage displaying known brain-homing peptide CAGALCY showed the highest fold-change over the G7 control phage (Fig. 5a), both in brain homing and selectivity. Also, BH1 and BH2 showed preferential accumulation in the brain (Fig. 5b-c) and LH1 showed tropism for the lungs (Fig. 5d).

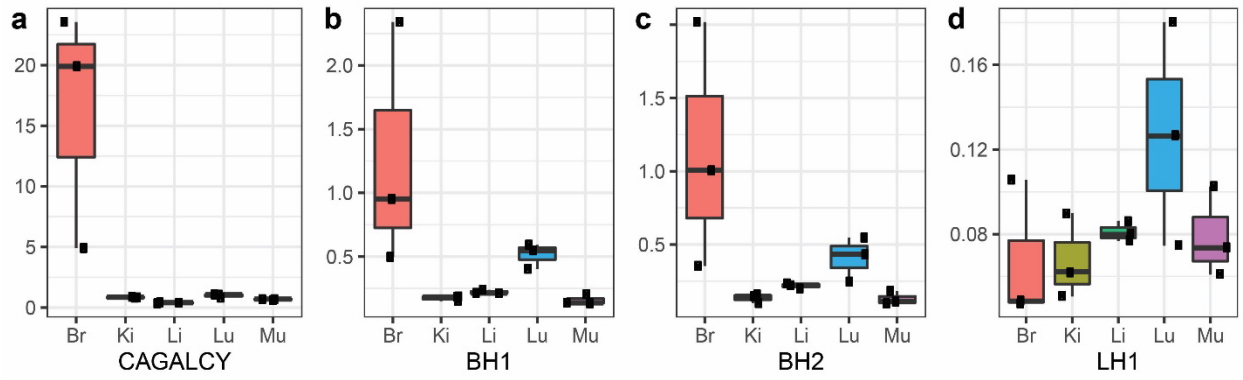


Fig. 5. Validation of homing specificity by in vivo peptide phage playoff and HTS.

*Individual peptide phages (G7, CAGALCY, BH1, BH2, LH1) were mixed at equimolar ratio and injected in BalbC mice (n=3). Results are expressed as fold change over G7 control phage. CAGALCY (a), BH1(b), BH2(c), LH1(d). Lower and upper hinges correspond to first and third quartiles, whiskers represent 1.5*IQR (inter-quartile range), middle line represents the median.*

In vivo homing of individual peptide phages

Biodistribution of peptide phages may be affected by other phages in the mix, with the bystander effect best characterized for tumor penetrating CendR peptides (33). Therefore, we decided to study biodistribution of individual candidate organ specific peptide phages. Individual phage clones displaying CAGALCY, BH1, BH2 and LH1 were i.v. injected in mice (n=3 for each peptide sequence). After 30 min circulation, the organs (brain, lung, liver, kidney, muscle) were collected and phage titer in each organ was determined. For each peptide phage, tissue distribution was expressed as fold change over control G7 phage (Fig. 6).

In the brain, CAGALCY-displaying phages showed ~ 65-fold overrepresentation compared to G7 phage, and the phages displaying newly identified brain homing peptides BH1 and BH2 showed ~20-fold overrepresentation. Second highest difference for BH1 and BH2 was observed in the lung. This was expected as both sequences did not show brain selectivity when compared to lung using differential binding approach.

Finally, we investigated whether BH1, BH2 or LH1 could be target-unrelated peptides, i.e. peptides that are enriched for reasons other than target recognition. TUPScan did not identify peptides BH1, BH2 or LH1 as target-unrelated peptide sequences. Furthermore, PhDFaster did

not detect peptides BH1, BH2, LH1 as sequences that could cause phages to be amplified faster than other clones. MimoScan confirmed that BH1, BH2 and LH1 have not been previously identified in other phage display datasets that are included in BDB.

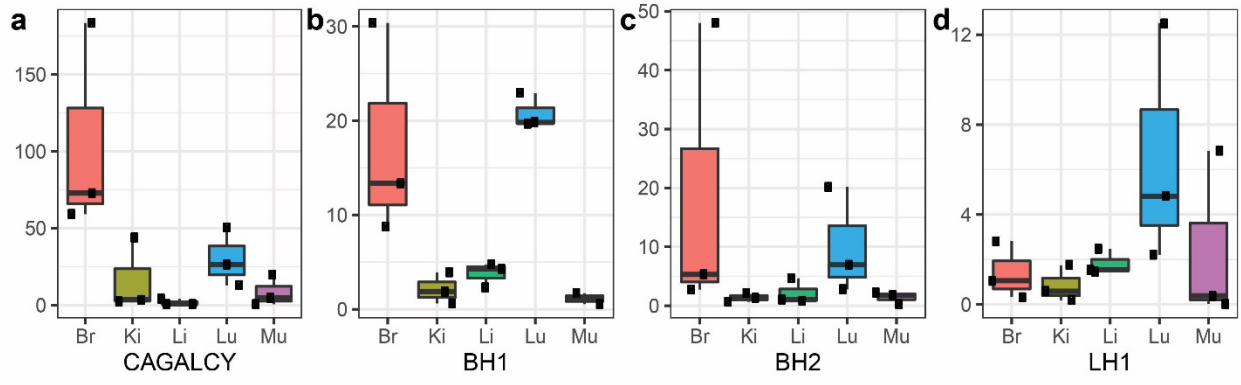


*Fig. 6. Im vivo biodistribution of individual peptide phages. Individual peptide- displaying phages (CAGALCY, BH1, BH2, LH1, G7) were injected in mice (n=3 for each peptide). Organs (brain, lung, liver, kidney, muscle) were collected and phages bound to the organ were titrated. Fold change (y axis) was calculated over the titer of G7 phage. CAGALCY (a) homes to brain. BH1(b) and BH2 (c) are selective towards brain when compared to other organs except lung. LH1 (d) shows selectivity towards lung. Lower and upper hinges correspond to first and third quartiles, whiskers represent 1.5*IQR (inter-quartile range), middle line represents the median.*

DISCUSSION

In vivo peptide phage display is a powerful technology that allows unbiased agnostic mapping of systemically accessible vascular heterogeneity in diseased and normal tissues and development of peptide affinity ligands for precision delivery of drugs, imaging agents and nanoparticles. However, the technology is laborious, and subject to biases and artefacts due to low stringency and reproducibility, and lack of validated tools for identification of the most avid binders in selected phage pools. Availability of powerful and affordable HTS platforms has made it feasible to determine homing peptide phage landscapes in target and control organs. In the current study, we developed an approach for identification of target-selective homing peptides based solely on in vivo display HTS data. Application of this approach enables faster and more reliable homing peptide discovery than provided by the current methods. In a study using a peptide T7 phage library with spiked-in brain homing phage demonstrated brain-specific differential binding of brain homing phage and resulted in identification of novel lung and brain selective homing peptides.

Our HTS-based screening addresses the main shortcoming in HTS-based phage biopanning – excessive reliance on enrichment of peptide phages at the target site. Enrichment alone does not provide full indication of selectivity of the peptide, as representation of peptide phages at the target site is also influenced by non-specific promiscuous interactions and peptide-dependent variation in the phage amplification rate (29). Our approach takes into account the statistically weighted enrichment for each peptide in the pool in the context of representation across target and nontarget sites. The method was inspired by published analysis of RNA sequencing data (34) and availability of computational tools such as *edgeR* that allow the use of quasi-likelihood F-tests and correct the results taking into account multiple comparisons (30)(35).

To validate our approach we decided to add to our starting library a phage displaying a published brain-selective positive control peptide, CAGALCY (19, 20). Adding a known homing peptide provides an internal control for the screen. In the screen reported here, the phage expressing the CAGALCY exhibited excellent brain specific accumulation, supporting the ability of the method to identify peptides of desired tropism. Remarkably, despite the low complexity of the peptide library we used, three novel lung selective peptides and five brain homing peptides were identified. Similar strategies for novel peptide identification using phage screens that take

advantage of statistical methods have already proven effective for HTS data in *in vitro* settings (29). HTS-based peptide differential profiling can be used for systematic identification of homing peptides and possibly other phage-displayed targeting ligands (e.g. single chain antibodies, affibodies). It should be possible to develop affinity ligands targeting vascular ZIP codes – systemically accessible vascular heterogeneity across normal and diseased tissues (36) following a similar process. Variations on the theme to be explored in follow-up studies include additional tissue fractionation steps for profiling of cell type and extracellular matrix specific peptides in normal organs and in pathological conditions such as tumors.

In summary, differential binding-based HTSin vivo phage display developed in the-- current study likely to render identification of vascular targeting peptides more reliable and consistent facilitating progress towards development of affinity guided smart drugs and contrast agents.

DATA AVAILABILITY

R script for differential binding analysis and generation of figures is available at https://github.com/KarlisPleiko/diff_phage.

FUNDING

This work was supported by the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012), by EMBO Installation grant #2344 (to T. Teesalu), European Research Council grant GLIOGUIDE (780915) from European Regional Development Fund (to T. Teesalu), Wellcome Trust International Fellowship WT095077MA (to T. Teesalu), and Norwegian-Estonian collaboration grant EMP181 (to T. Teesalu). We also acknowledge the support of Estonian Research Council (grants PRG230 and EAG79 to T. Teesalu). K.Pleiko was supported by PhD research scholarship administered by the University of Latvia Foundation and funded by “Mikrotīkls” Ltd. and ERASMUS scholarship.

ACKNOWLEDGMENTS

We thank Prof. Erkki Ruoslahti for reading and constructive criticism of the manuscript.

Author contributions: K.P., K.P., M.H. designed and carried out experiments, analyzed the data, composed the figures and wrote the manuscript. U.R. designed the differential binding

experiments. K.K. performed sequencing. T.T. conceived and supervised the project, analyzed the data and wrote the manuscript.

Graphic abstract, Fig.1 and Fig. 3 were created using BioRender.com.

REFERENCES

1. Smith,G.P. (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science (80-.).*, **228**, 1315–1317.

2. Henninot,A., Collins,J.C. and Nuss,J.M. (2018) The Current State of Peptide Drug Discovery: Back to the Future? *J. Med. Chem.*, **61**, 1382–1414.

3. Pasqualini,R. and Ruoslahti,E. (1996) Organ targeting in vivo using phage display peptide libraries. *Nature*, **380**, 364–366.

4. Arap,W., Pasqualini,R. and Ruoslahti,E. (1998) Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. *Science (80-.).*, **279**, 377–380.

5. Teesalu,T., Sugahara,K.N. and Ruoslahti,E. (2012) Mapping of vascular ZIP codes by phage display. In *Methods in Enzymology*. Academic Press Inc., Vol. 503, pp. 35–56.

6. Sugahara,K.N., Teesalu,T., Karmali,P.P., Kotamraju,V.R., Agemy,L., Girard,O.M., Hanahan,D., Mattrey,R.F. and Ruoslahti,E. (2009) Tissue-Penetrating Delivery of Compounds and Nanoparticles into Tumors. *Cancer Cell*, **16**, 510–520.

7. Teesalu,T., Sugahara,K.N., Kotamraju,V.R. and Ruoslahti,E. (2009) C-end rule peptides mediate neuropilin-1-dependent cell, vascular, and tissue penetration. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 16157–16162.

8. Ruoslahti,E. (2017) Tumor penetrating peptides for improved drug delivery. *Adv. Drug Deliv. Rev.*, **110–111**, 3–12.

9. Aumailley,M., Gurrath,M., Müller,G., Calvete,J., Timpl,R. and Kessler,H. (1991) Arg-Gly-Asp constrained with cyclic pentapeptides. *Febs*, **29**, 50–54.

10. Altmann,A., Sauter,M., Roesch,S., Mier,W., Warta,R., Debus,J., Dyckhoff,G., Herold-Mende,C. and Haberkorn,U. (2017) Identification of a novel ITGαvβ6-binding peptide

using protein separation and phage display. *Clin. Cancer Res.*, **23**, 4170–4180.

11. Pasqualini,R., Koivunen,E., Kain,R., Lahdenranta,J., Sakamoto,M., Stryhn,A., Ashmun,R.A., Shapiro,L.H., Arap,W. and Ruoslahti,E. (2000) Aminopeptidase N is a receptor for tumor-homing peptides and a target for inhibiting angiogenesis. *Cancer Res.*, **60**, 722–727.

12. Mann,A.P., Scodeller,P., Hussain,S., Braun,G.B., Mölder,T., Toome,K., Ambasudhan,R., Teesalu,T., Lipton,S.A. and Ruoslahti,E. (2017) Identification of a peptide recognizing cerebrovascular changes in mouse models of Alzheimer's disease. *Nat. Commun.*, **8**.

13. Han,Z., Zhou,Z., Shi,X., Wang,J., Wu,X., Sun,D., Chen,Y., Zhu,H., Magi-Galluzzi,C. and Lu,Z.R. (2015) EDB fibronectin specific peptide for prostate cancer targeting. *Bioconjug. Chem.*, **26**, 830–838.

14. Lingasamy,P., Tobi,A., Haugas,M., Hunt,H., Paiste,P., Asser,T., Rätsep,T., Kotamraju,V.R., Bjerkvig,R. and Teesalu,T. (2019) Bi-specific tenascin-C and fibronectin targeted peptide for solid tumor delivery. *Biomaterials*, **219**, 119373.

15. Lingasamy,P., Tobi,A., Kurm,K., Kopanchuk,S., Sudakov,A., Salumäe,M., Rätsep,T., Asser,T., Bjerkvig,R. and Teesalu,T. (2020) Tumor-penetrating peptide for systemic targeting of Tenascin-C. *Sci. Rep.*, **10**, 1–13.

16. Ikemoto,H., Lingasamy,P., Willmore,A.A., Hunt,H., Kurm,K., Tammik,O., Scodeller,P., Simón-gracia,L., Kotamraju,V.R., Lowy,A.M., *et al.* (2017) Hyaluronan-binding peptide for targeting peritoneal carcinomatosis. 10.1177/1010428317701628.

17. Urakami,T., Järvinen,T.A.H., Toba,M., Sawada,J., Ambalavanan,N., Mann,D., McMurtry,I., Oka,M., Ruoslahti,E. and Komatsu,M. (2011) Peptide-directed highly selective targeting of pulmonary arterial hypertension. *Am. J. Pathol.*, **178**, 2489–2495.

18. Scodeller,P., Simón-Gracia,L., Kopanchuk,S., Tobi,A., Kilk,K., Säälik,P., Kurm,K., Squadrito,M.L., Kotamraju,V.R., Rinken,A., *et al.* (2017) Precision Targeting of Tumor Macrophages with a CD206 Binding Peptide. *Sci. Rep.*, **7**.

19. Fan,X., Venegas,R., Fey,R., Van Der Heyde,H., Bernard,M.A., Lazarides,E. and Woods,C.M. (2007) An in vivo approach to structure activity relationship analysis of peptide ligands. *Pharm. Res.*, **24**, 868–879.

20. Toome,K., Willmore,A.M.A., Paiste,P., Tobi,A., Sugahara,K.N., Kirsimäe,K., Ruoslahti,E., Braun,G.B. and Teesalu,T. (2017) Ratiometric: In vivo auditioning of targeted silver nanoparticles. *Nanoscale*, **9**, 10094–10100.

21. Matochko,W.L., Cory Li,S., Tang,S.K.Y. and Derda,R. (2014) Prospective identification of parasitic sequences in phage display screens. *Nucleic Acids Res.*, **42**, 1784–1798.

22. Matochko,W.L., Chu,K., Jin,B., Lee,S.W., Whitesides,G.M. and Derda,R. (2012) Deep sequencing analysis of phage libraries using Illumina platform. *Methods*, **58**, 47–55.

23. Matochko,W.L., Ng,S., Jafari,M.R., Romaniuk,J., Tang,S.K.Y. and Derda,R. (2012) Uniform amplification of phage display libraries in monodisperse emulsions. *Methods*, **58**, 18–27.

24. He,B., Chen,H., Li,N. and Huang,J. (2019) Sarotup: A suite of tools for finding potential target-unrelated peptides from phage display data. *Int. J. Biol. Sci.*, **15**, 1452–1459.

25. He,B., Chen,H. and Huang,J. (2019) PHD7Faster 2.0: Predicting clones propagating faster from the Ph.D.-7 phage display library by coupling PseAAC and tripeptide composition. *PeerJ*, **2019**, 1–14.

26. Hurwitz,A.M., Huang,W., Estes,M.K., Atmar,R.L. and Palzkill,T. (2017) Deep sequencing of phage-displayed peptide libraries reveals sequence motif that detects norovirus. *Protein Eng. Des. Sel.*, **30**, 129–139.

27. Rebollo,I.R., Sabisz,M., Baeriswyl,V. and Heinis,C. (2014) Identification of target-binding peptide motifs by high-throughput sequencing of phage-selected peptides. *Nucleic Acids Res.*, **42**.

28. He,B., Jiang,L., Duan,Y., Chai,G., Fang,Y., Kang,J., Yu,M., Li,N., Tang,Z., Yao,P., *et al.* (2018) Biopanning data bank 2018: Hugging next generation phage display. *Database*, **2018**, 1–8.

29. Brinton,L.T., Bauknight,D.K., Dasa,S.S.K. and Kelly,K.A. (2016) PHASTpep: Analysis Software for Discovery of Cell-Selective Peptides via Phage Display and Next-Generation Sequencing. *PLoS One*, **11**, 1–22.

30. Chen,Y., Mccarthy,D., Robinson,M. and Smyth,G.K. (2014) edgeR: differential expression

analysis of digital gene expression data.

31. Conway,J.R., Lex,A. and Gehlenborg,N. (2017) UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics*, **33**, 2938–2940.

32. Wickham,H., Averick,M., Bryan,J., Chang,W., McGowan,L., François,R., Grolemund,G., Hayes,A., Henry,L., Hester,J., *et al.* (2019) Welcome to the Tidyverse. *J. Open Source Softw.*, **4**, 1686.

33. Sugahara,K.N., Teesalu,T., Karmali,P.P., Kotamraju,V.R., Agemy,L., Greenwald,D.R. and Ruoslahti,E. (2010) Coadministration of a Tumor-Penetrating Peptide Enhances the Efficacy of Cancer Drugs. *Science (80-.).*, **328**, 1031–1035.

34. Wang,Z., Gerstein,M. and Snyder,M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.

35. Lun,A.T.L., Chen,Y. and Smyth,G.K. (2016) It's DE-licious: A recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. In *Methods in Molecular Biology*. Humana Press Inc., Vol. 1418, pp. 391–416.

36. Arap,W., Kolonin,M.G., Trepel,M., Lahdenranta,J., Cardó-Vila,M., Giordano,R.J., Mintz,P.J., Ardelt,P.U., Yao,V.J., Vidal,C.I., *et al.* (2002) Steps toward mapping the human vasculature by phage display. *Nat. Med.*, **8**, 121–7.

FIG. LEGENDS

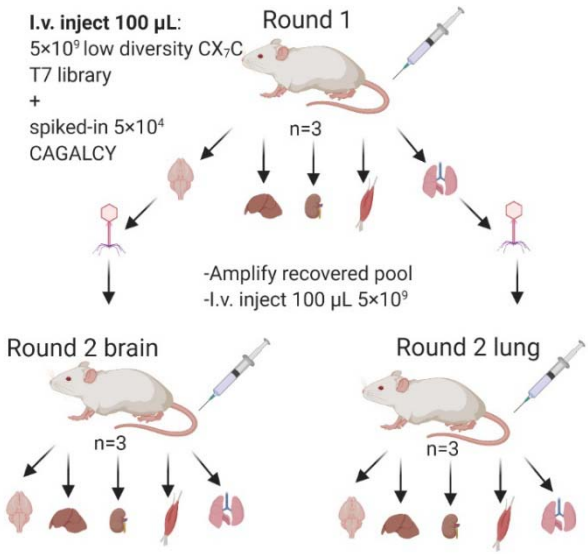


Fig. 1. Flow of in vivo phage biopanning. In round 1, 100 µL of 5×10^9 pfu collapsed CX_7C T7 phage library and 5×10^4 pfu of CAGALCY coding phage was injected in mice (n=3). Target organs (lung and brain) and control organs (kidney, liver, muscle) were collected and phage pools contained in target organs were amplified. For round 2, 5×10^9pfu of amplified libraries were injected in mice for further selection of lung (n=3) or brain (n=3) selective phages. Phage pools isolated from each organ at each round were subjected to HTS.

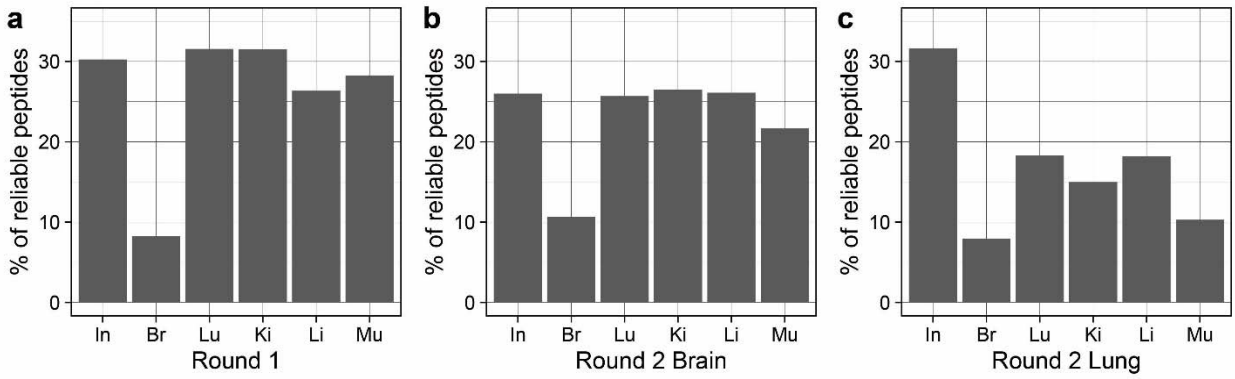


Fig. 2. Percentage of reliable peptides during in vivo selection rounds 1 and 2.

*The graphs represent percentage of reliable peptides (defined as being present in n=3 replicates for each sample) in mice dosed with naïve library **(a)**, round 1 brain-selected library **(b)**, and round 1 lung-selected library **(c)** . Input libraries represent technical variability between*

technical replicates (n=3) whereas organ-rescued libraries represent biological variability (n=3).

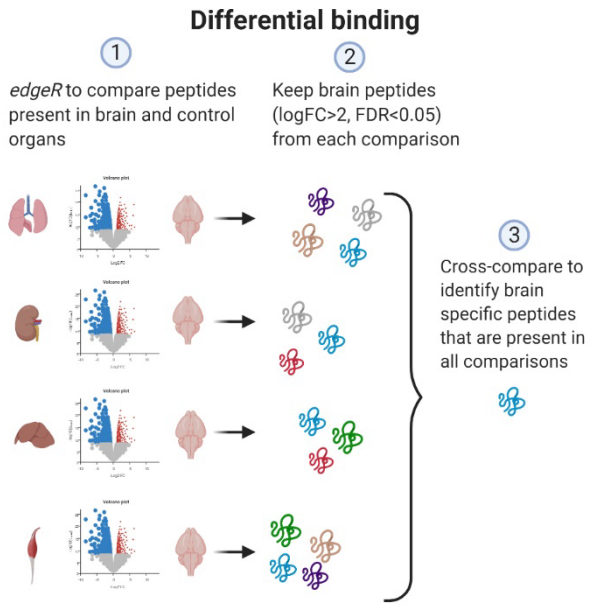


Fig. 3. Differential binding analysis.

HTS data from target organ (brain or lung) were compared to all control organs individually. Using edgeR, logFC > 2 and FDR <0.05 were applied as cutoff to identify peptides that are statistically signifficantly over-represented in brain. Peptides from each individual comparison were cross-compared to identify hits that appear in all comparisons as brain selective.

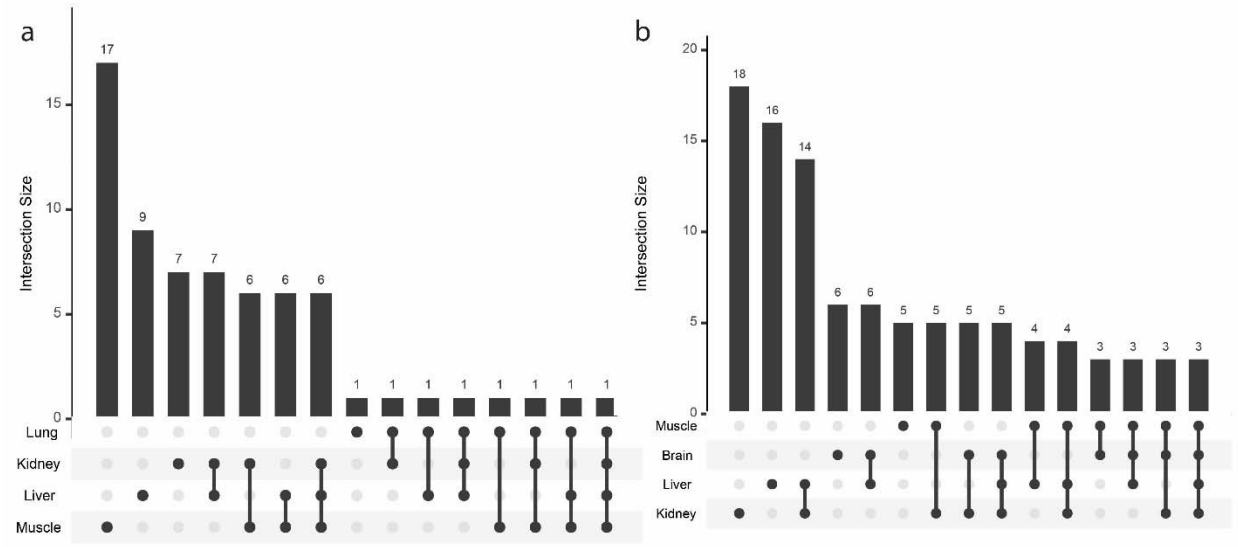


Fig. 4. Multiorgan comparison of selected phage pools to identify organ specific homing peptides.

Statistically significant (FDR < 0.05) brain homing (a) and lung homing (b) peptides compared against each control organ (logFC > 2, FDR < 0.05) and number of peptides that overlap between different comparisons. Only one peptide (CAGALCY) is brain selective and 3 peptides are lung selective when compared to all other control organs (Supplementary Table S5 and S6).

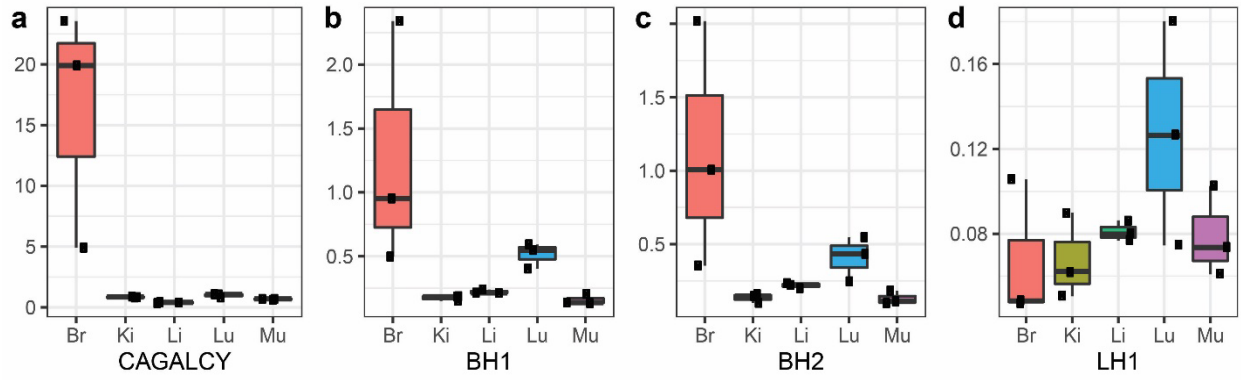


Fig. 5. Intra-animal validation using in vivo peptide phage playoff and HTS.

*Individual peptide phages (G7, CAGALCY, BH1, BH2, LH1) were mixed at equimolar ratio and injected in BalbC mice (n=3). Results were expressed as fold change over G7 control phage. CAGALCY (a), BH1(b), BH2(c), LH1(d). Lower and upper hinges correspond to first and third quartiles, whiskers represent 1.5*IQR (inter-quartile range), middle line represents the median.*

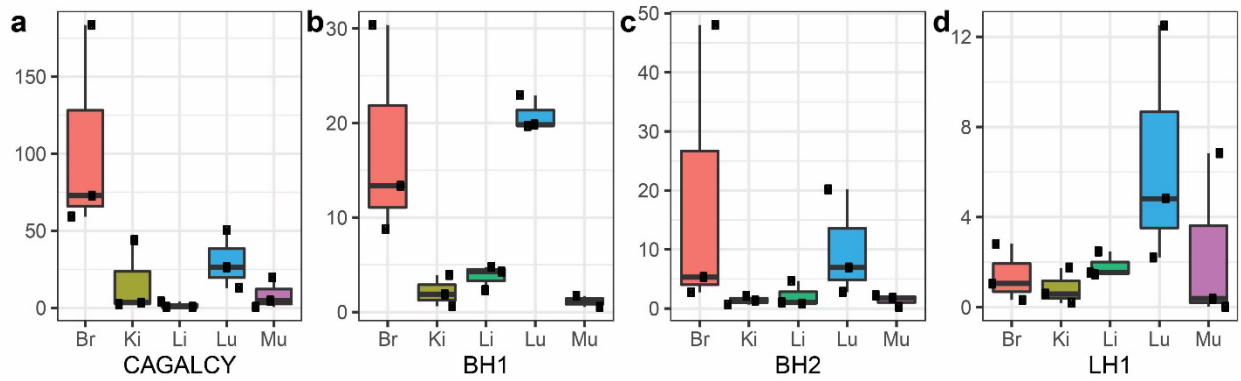


Fig. 6. Individual peptide-phage testing. Individual peptide displaying phages (CAGALCY, BH1, BH2, LH1, G7) were injected in mice (n=3 for each peptide). Organs (brain, lung, liver, kidney, muscle) were collected and phages bound to the organ were titered. Fold change (y axis) was

*calculated over the titer of G7 phage. CAGALCY (a) homes to brain. BH1(b) and BH2 (c) are selective towards brain when compared to other organs except lung. LH1 (d) shows selectivity towards lung. Lower and upper hinges correspond to first and third quartiles, whiskers represent 1.5*IQR (inter-quartile range), middle line represents the median.*